GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 10, 2002, 14:04:28 ; Search time 24.93 Seconds
(without alignments)
143.610 Million cell updates/sec

Title: US-09-445-803-14
Perfect score: 235
Sequence: 1 PLTNPIDAAIGGSADRNAEA.....DQIAAAMVLRGMAKDGQFAL 47

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	82	34.9	353	2	I40300	outer membrane lip
2	63	26.8	246	2	A43579	Vmp7 protein homol
3	63	26.8	353	2	F83570	hypothetical prote
4	59	25.1	367	2	F64202	probable GTP-bindi
5	58.5	24.9	637	2	T13432	gamma-glutamyltran
6	58	24.7	500	2	H86191	hypothetical prote
7	56	23.8	364	2	S11981	variable major pro
8	56	23.8	797	2	D86269	hypothetical prote
9	54.5	23.2	794	2	S59069	Z13 protein - mous
10	54	23.0	258	2	F71401	hypothetical prote
11	54	23.0	611	2	T44962	succinate dehydrog
12	54	23.0	686	2	B85153	hypothetical prote
13	54	23.0	915	2	T33030	hypothetical prote
14	54	23.0	941	2	T33032	hypothetical prote
15	53.5	22.8	2098	2	T25888	hypothetical prote
16	53	22.6	542	2	E84250	acetyl-CoA synthet
17	53	22.6	727	2	T10616	hypothetical prote
18	52.5	22.3	432	2	S48438	dnaJ protein homol
19	52.5	22.3	460	2	D84396	signal recognition
20	52.5	22.3	478	2	T26758	hypothetical prote
21	52.5	22.3	914	2	B75017	alanyl-tRNA synthe
22	52	22.1	666	2	D69366	conserved hypothet
23	52	22.1	1495	2	A85240	hypothetical prote
24	52	22.1	1495	2	T10649	hypothetical prote
25	51.5	21.9	245	2	T51773	acetoacetyl-CoA re
26	51.5	21.9	705	2	T19595	hypothetical prote
27	51	21.7	331	2	T18248	conserved hypothet
28	51	21.7	337	2	T46958	anthranilate phosp
29	51	21.7	363	2	F84360	threonine synthase
30	51	21.7	388	1	SYECSB	succinate--CoA lig
31	51	21.7	388	2	H85573	succinyl-CoA synth
32	51	21.7	391	2	T36311	probable lipase -
33	51	21.7	556	2	F82342	methyl-accepting c
34	51	21.7	649	2	T04005	probable protein k
35	50.5	21.5	284	2	S66520	tropomyosin, fast
36	50.5	21.5	694	2	T30725	probable abortive
37	50.5	21.5	1219	2	I61713	co-repressor prote
38	50.5	21.5	1229	2	A56068	co-repressor prote
39	50	21.3	77	2	F69219	conserved hypothet
40	50	21.3	247	2	T48878	proteasome psmA, a
41	50	21.3	338	2	T36785	hypothetical prote
42	50	21.3	362	2	S73454	probable GTP-bindi
43	50	21.3	412	2	B81176	hypothetical prote
44	50	21.3	630	2	C84453	probable selenium-
45	50	21.3	648	2	A48646	amine oxidase (cop

ALIGNMENTS

RESULT 1
I40300
outer membrane lipoprotein - Borrelia hermsii
C;Species: Borrelia hermsii
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 08-Oct-1999
C;Accession: I40300
R;Restrepo, B.I.; Kitten, T.; Carter, C.J.; Infante, D.; Barbour, A.G.
Mol. Microbiol. 6, 3299-3311, 1992
A;Title: Subtelomeric expression regions of Borrelia hermsii linear plasmids are high
A;Reference number: I40300; MUID:93133110
A;Accession: I40300
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-353 <RES>
A;Cross-references: GB:L04788; NID:g144023; PIDN:AAA22963.1; PID:g144024
C;Genetics:
A;Gene: vmp17

Query Match 34.9%; Score 82; DB 2; Length 353;
Best Local Similarity 40.0%; Pred. No. 0.01;
Matches 18; Conservative 8; Mismatches 19; Indels 0; Gaps 0;

Qy 3 TNPIDAAIGGSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
Db 237 TDAAEIAVAPAKDNKKEIKDGAKKDAVIAAGIALRAMAKNGTFSI 281

RESULT 2
A43579
Vmp7 protein homolog - Borrelia hermsii
C;Species: Borrelia hermsii
C;Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 23-Mar-1993
C;Accession: A43579
R;Barbour, A.G.; Carter, C.J.; Burman, N.; Freitag, C.S.; Garon, C.F.; Bergstroem, S.
Infect. Immun. 59, 390-397, 1991
A;Title: Tandem insertion sequence-like elements define the expression site for varia
A;Reference number: A43579; MUID:91099991
A;Accession: A43579
A;Molecule type: DNA
A;Residues: 1-246 <BAR>
A;Cross-references: GB:Z11876
C;Comment: This spirochete eludes the immune system to cause the disease relapsing fe
C;Keywords: membrane protein

Query Match 26.8%; Score 63; DB 2; Length 246;
Best Local Similarity 39.0%; Pred. No. 1.7;
Matches 16; Conservative 6; Mismatches 19; Indels 0; Gaps 0;

Qy 6 IDAAIGGSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFA 46

```
Db    138 LQAMVKNGANAAADAAKANAKDGTIAGAIALTVMAKGGKFA 178


RESULT   3
F83570
hypothetical protein PA0599 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: F83570
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathc
A;Reference number: A82950; MUID:20437337
A;Accession: F83570
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-353 <STO>
A;Cross-references: GB:AE004496; GB:AE004091; NID:g9946468; PIDN:AAG03988.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA0599


  Query Match             26.8%;  Score 63;  DB 2;  Length 353;
  Best Local Similarity   42.9%;  Pred. No. 2.4;
  Matches   12;  Conservative    5;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       1 PLTNPIDAAIGGSADRNAEAFDKMKKDD 28
           |:| | ::|   |||::| | |   | |
Db     137 PVTRPAESAASASADKDASAADSASKPD 164


RESULT   4
F64202
probable GTP-binding protein - Mycoplasma genitalium
C;Species: Mycoplasma genitalium
C;Date: 17-Nov-1995 #sequence_revision 17-Nov-1995 #text_change 02-Feb-2001
C;Accession: F64202; T09682
R;Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.;
M.; Fuhrmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J.
, C.A.; Venter, J.C.
Science 270, 397-403, 1995
A;Title: The minimal gene complement of Mycoplasma genitalium.
A;Reference number: A64200; MUID:96026346
A;Accession: F64202
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-367 <TIGR>
A;Cross-references: GB:U39680; GB:L43967; NID:g1045681; PID:g1045693; TIGR:MG024
A;Experimental source: strain G-37
R;Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.;
, M.; Fuhrmann, J.L.; Nguyen, D.T.; Utterback, T.; Saudek, D.M.; Phillips, C.A.; Merrick
.
submitted to the EMBL Data Library, October 1998
A;Reference number: Z16818
A;Accession: T09682
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-367 <FRA>
A;Cross-references: EMBL:U39681; NID:g3844626; PID:g3844633
C;Genetics:
A;Gene: MG024
A;Genetic code: SGC3
C;Superfamily: yeast probable purine nucleotide-binding protein YBR025c
C;Keywords: GTP; GTP binding; nucleotide binding; P-loop
F;8-15/Region: nucleotide-binding motif A (P-loop)
F;70-74/Region: GTP binding


  Query Match             25.1%;  Score 59;  DB 2;  Length 367;
  Best Local Similarity   34.8%;  Pred. No. 8;
  Matches   16;  Conservative    9;  Mismatches   15;  Indels    6;  Gaps    2;

Qy       2 LTNPIDAAIGGSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
           :|| |      |   | ||: ||: |:: :   :|| |: | ||  :
Db     139 ITNRI-----GKLKRKAESGDKIAKEEFVLLEIVLNGL-KQGQMPI 178


RESULT   5
T13432
gamma-glutamyltransferase homolog T17A13.30 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 18-Feb-2000
C;Accession: T13432
R;Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Ban
submitted to the Protein Sequence Database, July 1999
A;Reference number: Z17683
A;Accession: T13432
A;Molecule type: DNA
A;Residues: 1-637 <BEV>
A;Cross-references: EMBL:AL096692; GSPDB:GN00062; ATSP:T17A13.30
A;Experimental source: cultivar Columbia; BAC clone F17A13
C;Genetics:
A;Gene: ATSP:T17A13.30
A;Map position: 4
A;Introns: 43/1; 156/3; 497/3; 543/3
C;Superfamily: gamma-glutamyltransferase


  Query Match             24.9%;  Score 58.5;  DB 2;  Length 637;
  Best Local Similarity   28.1%;  Pred. No. 16;
  Matches   18;  Conservative    8;  Mismatches   19;  Indels   19;  Gaps    2;

Qy       2 LTNPIDAAIGGSA--------DRNAEAFD-----------KMKKDDQIAAAMVLRGMAKD 42
           : ||: : ||| :         |  |||||            | |:|  | ::    |
Db     117 VVNPMSSGIGGGSFLIVSSQKDSKAEAFDMRETAPLAASKDMYKNDASAKSLGALSMGVP 176

Qy      43 GQFA 46
           |: |
Db     177 GEIA 180


RESULT   6
H86191
hypothetical protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: H86191
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719
A;Accession: H86191
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-500 <STO>
A;Cross-references: GB:AE005172; NID:g6850304; PIDN:AAF29381.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1


  Query Match             24.7%;  Score 58;  DB 2;  Length 500;
  Best Local Similarity   37.5%;  Pred. No. 15;
  Matches   12;  Conservative    5;  Mismatches   15;  Indels    0;  Gaps    0;

Qy      12 GSADRNAEAFDKMKKDDQIAAAMVLRGMAKDG 43
           |  |  |: |||| : | |:   :: |  | |
```

```
Db    154 GQVDNAAKMFDKMPERDLISWTAMINGFVKKG 185


RESULT   7
S11981
variable major protein 21 - Borrelia hermsii
C;Species: Borrelia hermsii
C;Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 08-Oct-1999
C;Accession: S11981
R;Burman, N.; Bergstroem, S.; Restrepo, B.I.; Barbour, A.G.
Mol. Microbiol. 4, 1715-1726, 1990
A;Title: The variable antigens Vmp7 and Vmp21 of the relapsing fever bacterium Borrelia
A;Reference number: S11980; MUID:91171872
A;Accession: S11981
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-364 <BUR>
A;Cross-references: GB:M57256; EMBL:X53927; NID:g144025; PIDN:AAB59031.1; PID:g144026

  Query Match             23.8%;  Score 56;  DB 2;  Length 364;
  Best Local Similarity   29.8%;  Pred. No. 19;
  Matches   14;  Conservative   10;  Mismatches   15;  Indels    8;  Gaps    2;

Qy      3 TNPIDAAIGGSA---DRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFA 46
          |  |:: |:||:     :|| :    |   :|  : || : | |: |
Db    267 TTPLEFAVGGNGAHLSQNANS-----KAAAVAGGIALRSLVKGGKLA 308


RESULT   8
D86269
hypothetical protein AAF81289.1 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: D86269
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719
A;Accession: D86269
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-797 <STO>
A;Cross-references: GB:AE005172; NID:g8920567; PIDN:AAF81289.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1

  Query Match             23.8%;  Score 56;  DB 2;  Length 797;
  Best Local Similarity   34.2%;  Pred. No. 42;
  Matches   13;  Conservative    8;  Mismatches   13;  Indels    4;  Gaps    1;

Qy     12 GSADRNAEAFDKMKKD----DQIAAAMVLRGMAKDGQF 45
          |  |     |::|| |    | :| ::|: |: | |:|
Db    390 GRIDEALSLFNQMKADGLSPDLVAYSIVIHGLCKLGKF 427


RESULT   9
S59069
Z13 protein - mouse
C;Species: Mus musculus (house mouse)
C;Date: 27-Apr-1996 #sequence_revision 13-Mar-1997 #text_change 28-May-1999
C;Accession: S59069
R;Schulz, T.C.; Hopwood, B.; Rathjen, P.D.; Wells, J.R.E.
Biochem. J. 311, 219-224, 1995
A;Title: An unusual arrangement of 13 zinc fingers in the vertebrate gene Z13.
A;Reference number: S59069; MUID:96003919
A;Accession: S59069
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-794 <SCH>
A;Cross-references: GB:U14556; NID:g608136; PIDN:AAA85493.1; PID:g608137
C;Superfamily: POZ domain homology
F;10-108/Domain: POZ domain homology <POZ>

  Query Match             23.2%;  Score 54.5;  DB 2;  Length 794;
  Best Local Similarity   42.9%;  Pred. No. 64;
  Matches   18;  Conservative    6;  Mismatches   15;  Indels    3;  Gaps    3;

Qy      2 LTNPIDAAIGGSADRNA-EAFDKMKKDDQIAAAMVLR-GMAK 41
          |  | :  | ||| :| |  ||  ||:: || |: | | |:
Db    114 LAEP-SSTTGESADASAVEGGDKRAKDEKAAATMLSRLGQAR 154


RESULT  10
F71401
hypothetical protein - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
A;Variety: columbia
C;Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 05-Dec-1998
C;Accession: F71401
R;Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; D
 P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terryn, N.; G
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B
Nature 391, 485-488, 1998
A;Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomen
erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.;
C.; Chalwatzis, N.
A;Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis t
A;Reference number: A71400; MUID:98121113
A;Accession: F71401
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-258 <BEV>
A;Cross-references: GB:Z97335; NID:g2244747; PID:e326865; PID:g2244761
C;Genetics:
A;Map position: 4COP9-4G3845

  Query Match             23.0%;  Score 54;  DB 2;  Length 258;
  Best Local Similarity   36.6%;  Pred. No. 23;
  Matches   15;  Conservative    5;  Mismatches   19;  Indels    2;  Gaps    1;

Qy      1 PLTNPIDAAIG--GSADRNAEAFDKMKKDDQIAAAMVLRGM 39
          || | :    |  |:|    : ||:|   | || | ||  :
Db    107 PLANTLVNVYGKCGAASHALQVFDEMPHRDHIAWASVLTAL 147


RESULT  11
T44962
succinate dehydrogenase chain A homolog [imported] - Natronomonas pharaonis
N;Alternate names: flavoprotein homolog
C;Species: Natronomonas pharaonis
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 04-Mar-2000
C;Accession: T44962
R;Mattar, S.; Souquet, M.; Henrich, H.J.; Engelhard, M.
submitted to the EMBL Data Library, August 1996
A;Description: The first fully resolved primary structure of an archaeal succinate-de
A;Reference number: Z22881
A;Accession: T44962
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-611 <MAT>
A;Cross-references: EMBL:Y07709; PIDN:CAA68982.1
A;Experimental source: strain SP1 /28
C;Genetics:
A;Gene: sdhA
C;Superfamily: fumarate reductase flavoprotein; 3-oxosteroid 1-dehydrogenase homology
```

Query Match 23.0%; Score 54; DB 2; Length 611;
Best Local Similarity 35.1%; Pred. No. 57;
Matches 13; Conservative 6; Mismatches 18; Indels 0; Gaps 0;

```
Qy       6 IDAAIGGSADRNAEAFDKMKKDDQIAAAMVLRGMAKD 42
           |:|||    |    |:| ||  | :  |  :  :|:|
Db      51 INAAIRDGDDWELHAYDTMKGSDYLGDAPAIETLAQD 87
```

RESULT 12
B85153
hypothetical protein AT4g14050 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C;Accession: B85153
R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sprin
Nature 402, 769-777, 1999
A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A;Reference number: A85001; MUID:20083488
A;Accession: B85153
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-686 <STO>
A;Cross-references: GB:NC_001268; NID:g7268109; PIDN:CAB78447.1; GSPDB:GN00140
C;Genetics:
A;Gene: AT4g14050
A;Map position: 4
C;Superfamily: Arabidopsis thaliana hypothetical protein At2g41080

Query Match 23.0%; Score 54; DB 2; Length 686;
Best Local Similarity 36.6%; Pred. No. 64;
Matches 15; Conservative 5; Mismatches 19; Indels 2; Gaps 1;

```
Qy       1 PLTNPIDAAIG--GSADRNAEAFDKMKKDDQIAAAMVLRGM 39
           || | :    |  |:|    : ||:|   | || | ||  :
Db     107 PLANTLVNVYGKCGAASHALQVFDEMPHRDHIAWASVLTAL 147
```

RESULT 13
T33030
hypothetical protein M70.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T33030
R;Miller, N.; Wamsley, P.; Twyman, B.
submitted to the EMBL Data Library, February 1998
A;Description: The sequence of C. elegans cosmid M70.
A;Reference number: Z21266
A;Accession: T33030
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-915 <MIL>
A;Cross-references: EMBL:AF047661; PIDN:AAC04437.1; GSPDB:GN00022; CESP:M70.3
A;Experimental source: strain Bristol N2; clone M70
C;Genetics:
A;Gene: CESP:M70.3
A;Map position: 4
A;Introns: 20/1; 49/1; 78/1; 98/1; 793/2; 891/3

Query Match 23.0%; Score 54; DB 2; Length 915;
Best Local Similarity 45.5%; Pred. No. 86;
Matches 15; Conservative 2; Mismatches 10; Indels 6; Gaps 1;

```
Qy       5 PIDAA------IGGSADRNAEAFDKMKKDDQIA 31
           | | |      |||   :: | ||| | || ||
Db     130 PSDEAITDLLNIGGVKLKDLETFDKTKVDDVIA 162
```

RESULT 14
T33032
hypothetical protein M70.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T33032
R;Miller, N.; Wamsley, P.; Twyman, B.
submitted to the EMBL Data Library, February 1998
A;Description: The sequence of C. elegans cosmid M70.
A;Reference number: Z21266
A;Accession: T33032
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-941 <MIL>
A;Cross-references: EMBL:AF047661; PIDN:AAC04435.1; GSPDB:GN00022; CESP:M70.1
A;Experimental source: strain Bristol N2; clone M70
C;Genetics:
A;Gene: CESP:M70.1
A;Map position: 4
A;Introns: 20/1; 40/1; 60/1; 80/1; 774/2; 892/3

Query Match 23.0%; Score 54; DB 2; Length 941;
Best Local Similarity 45.5%; Pred. No. 88;
Matches 15; Conservative 2; Mismatches 10; Indels 6; Gaps 1;

```
Qy       5 PIDAA------IGGSADRNAEAFDKMKKDDQIA 31
           | | |      |||   :: | ||| | || ||
Db     112 PSDEAITDLLNIGGVKLKDLETFDKTKVDDVIA 144
```

RESULT 15
T25888
hypothetical protein T10H10.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 08-Sep-2000
C;Accession: T25888
R;Nelson, J.; Langston, Y.
submitted to the EMBL Data Library, December 1996
A;Description: The sequence of C. elegans cosmid T10H10.
A;Reference number: Z20105
A;Accession: T25888
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2098 <NEL>
A;Cross-references: EMBL:U80848; PIDN:AAB37988.1; GSPDB:GN00028; CESP:T10H10.1
A;Experimental source: strain Bristol N2; clone T10H10
C;Genetics:
A;Gene: CESP:T10H10.1
A;Map position: X
A;Introns: 6/3; 43/3; 92/3; 154/2; 226/1; 280/3; 417/2; 472/3; 812/3; 853/3; 887/3; 9
/3
C;Superfamily: myosin motor domain homology
F;65-720/Domain: myosin motor domain homology <MMO>

Query Match 22.8%; Score 53.5; DB 2; Length 2098;
Best Local Similarity 48.1%; Pred. No. 2.3e+02;
Matches 13; Conservative 5; Mismatches 8; Indels 1; Gaps 1;

```
Qy      21 FDKMKKDDQI-AAAMVLRGMAKDGQFA 46
           : |  |:| |  ||::|| | |||: |
Db    1903 YHKTTKNDVIELAALILRSMTKDGKNA 1929
```

Search completed: January 10, 2002, 14:04:30
Job time: 124 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 10, 2002, 14:09:57 ; Search time 15.37 Seconds
(without alignments)
112.118 Million cell updates/sec

Title: US-09-445-803-14
Perfect score: 235
Sequence: 1 PLTNPIDAAIGGSADRNAEA.....DQIAAAMVLRGMAKDGQFAL 47

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	82	34.9	353	1	VM17_BORHE	P32777 borrelia he
2	59	25.1	367	1	Y024_MYCGE	P47270 mycoplasma
3	56	23.8	364	1	VM21_BORHE	P21875 borrelia he
4	54.5	23.2	794	1	Z151_MOUSE	Q60821 mus musculu
5	53.5	22.8	580	1	STCD_RHIME	O87278 rhizobium m
6	52.5	22.3	432	1	YIS4_YEAST	P40564 saccharomyc
7	52.5	22.3	460	1	SR54_HALN1	Q9hmn5 halobacteri
8	52.5	22.3	914	1	SYA_PYRAB	Q9uy36 pyrococcus
9	51	21.7	260	1	PCNA_SARCR	O16852 sarcophaga
10	51	21.7	337	1	TRPD_BRAJA	P94326 bradyrhizob
11	51	21.7	388	1	SUCC_ECOLI	P07460 escherichia
12	51	21.7	401	1	P39_BRUAB	O06875 brucella ab
13	51	21.7	804	1	SYM_ORYSA	Q9ztsl oryza sativ
14	50	21.3	247	1	PSMA_METTE	Q59565 methanosarc
15	50	21.3	362	1	Y024_MYCPN	P75088 mycoplasma
16	50	21.3	648	1	AMO1_ARTS1	Q07121 arthrobacte
17	50	21.3	648	1	AMO2_ARTS1	Q07123 arthrobacte
18	50	21.3	754	1	RRP1_INCJJ	P19703 influenza c
19	50	21.3	1011	1	UBAL_HUMAN	P41226 homo sapien
20	49.5	21.1	327	1	GBLP_ARATH	O24456 arabidopsis
21	49.5	21.1	337	1	TRPD_HALVO	P52562 halobacteri
22	49.5	21.1	1147	1	KIN2_YEAST	P13186 saccharomyc
23	49	20.9	234	1	PYRE_METAN	O42767 metarhizium
24	49	20.9	389	1	PYRD_DROME	P32748 drosophila
25	49	20.9	495	1	RHSB_RHIME	Q9z3r1 rhizobium m
26	49	20.9	560	1	TATR_NPVOP	P22114 orgyia pseu
27	49	20.9	807	1	GYRB_RICPR	Q9zcx2 rickettsia
28	49	20.9	941	1	CDAC_HUMAN	Q9un75 homo sapien
29	49	20.9	1097	1	CCT_DROME	O96433 drosophila
30	48.5	20.6	98	1	FER1_SYNP7	P06517 synechococc
31	48.5	20.6	349	1	PHOL_MYCLE	Q49751 mycobacteri
32	48.5	20.6	369	1	VM07_BORHE	P21876 borrelia he
33	48.5	20.6	504	1	SR54_HUMAN	P13624 homo sapien
34	48.5	20.6	504	1	SR54_MOUSE	P14576 mus musculu
35	48	20.4	251	1	YC52_METJA	Q58648 methanococc
36	48	20.4	293	1	RK4_SPIOL	O49937 spinacia ol
37	48	20.4	296	1	BLAC_BACUN	P30898 bacteroides
38	48	20.4	334	1	GBLP_ORYSA	P49027 oryza sativ
39	48	20.4	344	1	CHEB_THEMA	Q9wyn9 thermotoga
40	48	20.4	344	1	ETFA_YEAST	Q12480 saccharomyc
41	48	20.4	390	1	PGK_BUCAI	P57525 buchnera ap
42	48	20.4	427	1	RU17_ARATH	Q42404 arabidopsis
43	48	20.4	457	1	YE07_MYCTU	P71675 mycobacteri
44	48	20.4	594	1	YKA5_YEAST	P36106 saccharomyc
45	48	20.4	1422	1	RPOM_NEUCR	P38671 neurospora

ALIGNMENTS

RESULT 1
VM17_BORHE
ID VM17_BORHE STANDARD; PRT; 353 AA.
AC P32777;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE VARIABLE MAJOR OUTER MEMBRANE LIPOPROTEIN 17 PRECURSOR.
GN VMP17.
OS Borrelia hermsii.
OG Plasmid.
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=140;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SSP. HS1 SEROTYPE 17;
RX MEDLINE=93133110; PubMed=1484486;
RA Restrepo B.I., Kitten T., Carter C.J., Infante D., Barbour A.G.;
RT "Subtelomeric expression regions of Borrelia hermsii linear plasmids
RT are highly polymorphic.";
RL Mol. Microbiol. 6:3299-3311(1992).
CC -!- FUNCTION: SERVES TO AVOID THE HOST IMMUNE RESPONSE BY CHANGING
CC FROM ONE SURFACE EXPOSED VMP TO ANOTHER.
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID
CC ANCHOR.
CC ---

CC ---
DR EMBL; L04788; AAA22963.1; -.
DR InterPro; IPR000680; Borrelia_lipo.
DR InterPro; IPR001800; Lipoprotein_6.
DR Pfam; PF00921; Lipoprotein_2; 1.
DR Pfam; PF01441; Lipoprotein_6; 1.
DR ProDom; PD001149; Lipoprotein_6; 1.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Outer membrane; Lipoprotein; Signal; Plasmid.
FT SIGNAL 1 18 PROBABLE.
FT CHAIN 19 353 VARIABLE MAJOR OUTER MEMBRANE
FT LIPOPROTEIN 17.
FT LIPID 19 19 N-ACYL DIGLYCERIDE (PROBABLE).
SQ SEQUENCE 353 AA; 36460 MW; 88FAB9B8C619AE91 CRC64;

Query Match 34.9%; Score 82; DB 1; Length 353;
Best Local Similarity 40.0%; Pred. No. 0.0045;
Matches 18; Conservative 8; Mismatches 19; Indels 0; Gaps 0;

```
Qy       3 TNPIDAAIGGSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
           |:  : |:  : |   |  |  |||  ||| : || |||:| |::
Db     237 TDAAEIAVAPAKDNKKEIKDGAKKDAVIAAGIALRAMAKNGTFSI 281
```

```
RESULT   2
Y024_MYCGE
ID   Y024_MYCGE     STANDARD;      PRT;   367 AA.
AC   P47270;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PROBABLE GTP-BINDING PROTEIN MG024.
GN   MG024.
OS   Mycoplasma genitalium.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2097;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 33530 / G-37;
RX   MEDLINE=96026346; PubMed=7569993;
RA   Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA   Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA   Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA   Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA   Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA   Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT   "The minimal gene complement of Mycoplasma genitalium.";
RL   Science 270:397-403(1995).
CC   -!- SIMILARITY: BELONGS TO THE YCHF FAMILY OF GTP-BINDING PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U39681; AAC71240.1; -.
DR   TIGR; MG024; -.
DR   InterPro; IPR000765; GTP1_OBG.
DR   Pfam; PF01018; GTP1_OBG; 1.
DR   PRINTS; PR00326; GTP1OBG.
KW   GTP-binding; Complete proteome.
FT   NP_BIND       8     15       GTP (POTENTIAL).
FT   NP_BIND      71     75       GTP (POTENTIAL).
SQ   SEQUENCE   367 AA;  41198 MW;  98AFCFADDE6C9BDC CRC64;

  Query Match           25.1%;  Score 59;  DB 1;  Length 367;
  Best Local Similarity   34.8%;  Pred. No. 3.6;
  Matches   16;  Conservative    9;  Mismatches   15;  Indels    6;  Gaps    2;

Qy       2 LTNPIDAAIGGSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
           :|| |     |   | ||: ||: |:: :   :|| |: | ||  :
Db     139 ITNRI-----GKLKRKAESGDKIAKEEFVLLEIVLNGL-KQGQMPI 178


RESULT   3
VM21_BORHE
ID   VM21_BORHE     STANDARD;      PRT;   364 AA.
AC   P21875;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   01-FEB-1994 (Rel. 28, Last annotation update)
DE   VARIABLE MAJOR OUTER MEMBRANE LIPOPROTEIN 21 PRECURSOR.
GN   VMP21.
OS   Borrelia hermsii.
OG   Plasmid.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=140;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SSP. HS1 SEROTYPE 21;
RX   MEDLINE=91171872; PubMed=1706456;
RA   Burman N., Bergstroem S., Restrepo B.I., Barbour A.G.;
RT   "The variable antigens Vmp7 and Vmp21 of the relapsing fever
RT   bacterium Borrelia hermsii are structurally analogous to the VSG
RT   proteins of the African trypanosome.";
RL   Mol. Microbiol. 4:1715-1726(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SSP. HS1 SEROTYPE 21;
RX   MEDLINE=93133110; PubMed=1484486;
RA   Restrepo B.I., Kitten T., Carter C.J., Infante D., Barbour A.G.;
RT   "Subtelomeric expression regions of Borrelia hermsii linear plasmids
RT   are highly polymorphic.";
RL   Mol. Microbiol. 6:3299-3311(1992).
RN   [3]
RP   SEQUENCE OF 55-75; 186-208 AND 245-259.
RX   MEDLINE=85236116; PubMed=2409197;
RA   Barstad P.A., Coligan J.E., Raum M.G., Barbour A.G.;
RT   "Variable major proteins of Borrelia hermsii. Epitope mapping and
RT   partial sequence analysis of CNBr peptides.";
RL   J. Exp. Med. 161:1302-1314(1985).
CC   -!- FUNCTION: SERVES TO AVOID THE HOST IMMUNE RESPONSE BY CHANGING
CC       FROM ONE SURFACE EXPOSED VMP TO ANOTHER.
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID
CC       ANCHOR.
CC   -!- SIMILARITY: STRONG, TO VMP7 AND VMP25.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M57256; AAB59031.1; -.
DR   PIR; S11981; S11981.
DR   InterPro; IPR000680; Borrelia_lipo.
DR   Pfam; PF00921; Lipoprotein_2; 1.
DR   PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW   Outer membrane; Lipoprotein; Signal; Plasmid.
FT   SIGNAL        1     26       PROBABLE.
FT   CHAIN        27    364       VARIABLE MAJOR OUTER MEMBRANE
FT                                LIPOPROTEIN 21.
FT   LIPID        27     27       N-ACYL DIGLYCERIDE (PROBABLE).
SQ   SEQUENCE   364 AA;  37084 MW;  16598B639EE63776 CRC64;

  Query Match           23.8%;  Score 56;  DB 1;  Length 364;
  Best Local Similarity   29.8%;  Pred. No. 8.5;
  Matches   14;  Conservative   10;  Mismatches   15;  Indels    8;  Gaps    2;

Qy       3 TNPIDAAIGGSA---DRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFA 46
           | |:: |:||:     :|| :     |   :|  : || : | |: |
Db     267 TTPLEFAVGGNGAHLSQNANS-----KAAAVAGGIALRSLVKGGKLA 308


RESULT   4
Z151_MOUSE
ID   Z151_MOUSE     STANDARD;      PRT;   794 AA.
AC   Q60821; Q60699;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   ZINC FINGER PROTEIN 151 (POLYOMAVIRUS LATE INITIATOR PROMOTER BINDING
DE   PROTEIN) (LP-1) (ZINC FINGER PROTEIN Z13).
GN   ZNF151 OR ZFP100.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RA   Rapp L., Carmichael G.G.;
RL   Submitted (APR-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CBA; TISSUE=Kidney;
RX   MEDLINE=96003919; PubMed=7575457;
RA   Schulz T.C., Hopwood B., Rathjen P.D., Wells J.R.E.;
RT   "An unusual arrangement of 13 zinc fingers in the vertebrate gene
RT   Z13.";
RL   Biochem. J. 311:219-224(1995).
CC   -!- FUNCTION: MAY FUNCTION AS A HOUSEKEEPING DNA-BINDING PROTEIN THAT
CC       REGULATES THE EXPRESSION OF SPECIFIC GENES.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC   -!- TISSUE SPECIFICITY: FOUND IN ALL THE EMBRYONIC AND ADULT TISSUES
CC       EXAMINED.
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   -!- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U22396; AAA64848.1; -.
DR   EMBL; U14556; AAA85493.1; -.
DR   HSSP; P08047; 1SP2.
DR   MGD; MGI:107410; Zfp100.
DR   InterPro; IPR000210; BTB_POZ.
DR   InterPro; IPR000822; Znf-C2H2.
DR   Pfam; PF00651; BTB; 1.
DR   Pfam; PF00096; zf-C2H2; 13.
DR   PRINTS; PR00048; ZINCFINGER.
DR   SMART; SM00225; BTB; 1.
DR   SMART; SM00355; ZnF_C2H2; 13.
DR   PROSITE; PS50097; BTB; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 13.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 13.
KW   Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW   Nuclear protein; Repeat.
FT   DOMAIN        1    104       BTB.
FT   DOMAIN      297    730       ZINC FINGERS.
FT   ZN_FING     297    319       C2H2-TYPE.
FT   ZN_FING     325    347       C2H2-TYPE.
FT   ZN_FING     353    375       C2H2-TYPE.
FT   ZN_FING     381    403       C2H2-TYPE.
FT   ZN_FING     409    431       C2H2-TYPE.
FT   ZN_FING     437    459       C2H2-TYPE.
FT   ZN_FING     465    487       C2H2-TYPE.
FT   ZN_FING     493    515       C2H2-TYPE.
FT   ZN_FING     519    543       C2H2-TYPE.
FT   ZN_FING     549    571       C2H2-TYPE.
FT   ZN_FING     577    599       C2H2-TYPE.
FT   ZN_FING     605    628       C2H2-TYPE.
FT   ZN_FING     708    730       C2H2-TYPE.
FT   CONFLICT    507    507       G -> A (IN REF. 2).
FT   CONFLICT    573    573       N -> K (IN REF. 2).
SQ   SEQUENCE   794 AA;  86664 MW;  FFF88E56EDEBF7ED CRC64;
```

```
  Query Match             23.2%;  Score 54.5;  DB 1;  Length 794;
  Best Local Similarity   42.9%;  Pred. No. 30;
  Matches   18;  Conservative    6;  Mismatches   15;  Indels    3;  Gaps    3;

Qy       2 LTNPIDAAIGGSADRNA-EAFDKMKKDDQIAAAMVLR-GMAK 41
           |  |  :  | |||  :| |  ||  ||:: || |: | | |:
Db     114 LAEP-SSTTGESADASAVEGGDKRAKDEKAAATMLSRLGQAR 154
```

```
RESULT   5
STCD_RHIME
ID   STCD_RHIME     STANDARD;      PRT;   580 AA.
AC   O87278;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PROBABLE N-METHYLPROLINE DEMETHYLASE (EC 1.-.-.-) (STACHYDRINE
DE   UTILIZATION PROTEIN STCD).
GN   STCD.
OS   Rhizobium meliloti (Sinorhizobium meliloti).
OG   Plasmid pSymB (megaplasmid 2).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Sinorhizobium.
OX   NCBI_TaxID=382;
RN   [1]
RP   SEQUENCE FROM N.A., AND FUNCTION.
RC   STRAIN=1021;
RX   MEDLINE=98432824; PubMed=9758825;
RA   Phillips D.A., Sande E.S., Vriezen J.A.C., de Bruijn F.J.,
RA   Le Rudulier D., Joseph C.M.;
RT   "A new genetic locus in Sinorhizobium meliloti is involved in
RT   stachydrine utilization.";
RL   Appl. Environ. Microbiol. 64:3954-3960(1998).
CC   -!- FUNCTION: POSSIBLE NADH-DEPENDENT OXIDASE, FUNCTIONS AS A
CC       DEMETHYLASE THAT CONVERTS N-METHYLPROLINE TO PROLINE.
CC   -!- COFACTOR: FAD (POTENTIAL).
CC   -!- PATHWAY: SECOND STEP IN STACHYDRINE DEGRADATION.
CC   -!- INDUCTION: BY STACHYDRINE.
CC   -!- SIMILARITY: TO THE OYE FAMILY OF NADH-DEPENDENT FLAVIN
CC       OXIDOREDUCTASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF016307; AAC63907.1; -.
DR   InterPro; IPR000205; NAD_binding.
DR   InterPro; IPR001155; Oxidored_FMN.
DR   Pfam; PF00724; oxidored_FMN; 1.
KW   Oxidoreductase; NAD; FAD; Flavoprotein; Plasmid.
SQ   SEQUENCE   580 AA;  63938 MW;  40BDA5DC5F096F3D CRC64;
```

```
  Query Match             22.8%;  Score 53.5;  DB 1;  Length 580;
  Best Local Similarity   50.0%;  Pred. No. 29;
  Matches   13;  Conservative    2;  Mismatches   10;  Indels    1;  Gaps    1;

Qy       1 PLTNPIDAAIGGSAD-RNAEAFDKMK 25
           |||| :|   ||| | |    || :|
Db     179 PLTNELDGPYGGSLDNRMRFCFDVLK 204
```

```
RESULT   6
YIS4_YEAST
ID   YIS4_YEAST     STANDARD;      PRT;   432 AA.
AC   P40564;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   HYPOTHETICAL 48.6 KDA PROTEIN IN BET1-PAN1 INTERGENIC REGION.
GN   YIR004W OR YIB4W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
```

```
RA   Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA   Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA   Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA   Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA   Walsh S.V., Whitehead S.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE=95282515; PubMed=7762303;
RA   Voss H., Tamames J., Teodoru C., Valencia A., Sensen C., Wiemann S.,
RA   Schwager C., Zimmermann J., Sander C., Ansorge W.;
RT   "Nucleotide sequence and analysis of the centromeric region of yeast
RT   chromosome IX.";
RL   Yeast 11:61-78(1995).
CC   -!- SIMILARITY: CONTAINS 1 J DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z38062; CAA86206.1; -.
DR   EMBL; X79743; -; NOT_ANNOTATED_CDS.
DR   PIR; S48438; S48438.
DR   HSSP; P08622; 1XBL.
DR   SGD; S0001443; YIR004W.
DR   InterPro; IPR001623; DnaJ_N.
DR   Pfam; PF00226; DnaJ; 1.
DR   SMART; SM00271; DnaJ; 1.
DR   PROSITE; PS00636; DNAJ_1; 1.
DR   PROSITE; PS50076; DNAJ_2; 1.
KW   Hypothetical protein; Chaperone.
FT   DOMAIN        4     73       J-DOMAIN.
SQ   SEQUENCE   432 AA;  48574 MW;  8127D6886BC78B96 CRC64;


  Query Match             22.3%;  Score 52.5;  DB 1;  Length 432;
  Best Local Similarity   29.3%;  Pred. No. 28;
  Matches   12;  Conservative    8;  Mismatches   10;  Indels   11;  Gaps    1;

Qy       3 TNPIDAAIGGSADR----------NAEAFDKMKKDDQIAA 32
           || :|||:| : ::           |    |  ||::|: |
Db     145 TNAVDAALGNTNEKDDKNKARTTSGNLTVHDGNKKNEQVGA 185


RESULT   7
SR54_HALN1
ID   SR54_HALN1     STANDARD;      PRT;   460 AA.
AC   Q9HMN5;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   SIGNAL RECOGNITION 54 KDA PROTEIN (SRP54).
GN   SRP54 OR VNG2459G.
OS   Halobacterium sp. (strain NRC-1).
OC   Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC   Halobacterium.
OX   NCBI_TaxID=64091;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20504483; PubMed=11016950;
RA   Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA   Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA   Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA   Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA   Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA   Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA   Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA   Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT   "Genome sequence of Halobacterium species NRC-1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
CC   -!- FUNCTION: BINDS TO THE SIGNAL SEQUENCE OF PRESECRETORY PROTEIN
CC       WHEN THEY EMERGE FROM THE RIBOSOMES (BY SIMILARITY).
CC   -!- SUBUNIT: ARCHAEAL SIGNAL RECOGNITION PARTICLE CONSISTS OF A 7S RNA
CC       MOLECULE OF 300 NUCLEOTIDES AND TWO PROTEIN SUBUNITS: SRP54 AND
CC       SRP19 (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC   -!- DOMAIN: HAS A TWO DOMAIN STRUCTURE: THE G-DOMAIN BINDS GTP; THE
CC       M-DOMAIN BINDS THE 7S RNA IN PRESENCE OF SRP19 AND ALSO BINDS THE
CC       SIGNAL SEQUENCE (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE005125; AAG20536.1; -.
DR   InterPro; IPR000897; SRP54.
DR   Pfam; PF00448; SRP54; 1.
DR   ProDom; PD000819; SRP54; 1.
DR   PROSITE; PS00300; SRP54; FALSE_NEG.
KW   Signal recognition particle; GTP-binding; RNA-binding;
KW   Complete proteome.
FT   DOMAIN        1    289       G-DOMAIN (BY SIMILARITY).
FT   DOMAIN      290    460       M-DOMAIN (BY SIMILARITY).
FT   NP_BIND     104    111       GTP (BY SIMILARITY).
FT   NP_BIND     184    188       GTP (BY SIMILARITY).
FT   NP_BIND     242    245       GTP (BY SIMILARITY).
FT   DOMAIN      449    458       POLY-GLY.
SQ   SEQUENCE   460 AA;  50149 MW;  8361B782F651352E CRC64;


  Query Match             22.3%;  Score 52.5;  DB 1;  Length 460;
  Best Local Similarity   42.1%;  Pred. No. 30;
  Matches   16;  Conservative    3;  Mismatches   18;  Indels    1;  Gaps    1;

Qy       6 IDAAIGGSADRNAEAFDKMKKDDQIAAAMVLRGMAKDG 43
           :|||||  | : |  ||      | :|    | | || |
Db     215 LDAAIGQGAKQQARQFDASIGIDGVAITK-LDGTAKGG 251


RESULT   8
SYA_PYRAB
ID   SYA_PYRAB      STANDARD;      PRT;   914 AA.
AC   Q9UY36;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   ALANYL-TRNA SYNTHETASE (EC 6.1.1.7) (ALANINE--TRNA LIGASE) (ALARS).
GN   ALAS OR PAB1245.
OS   Pyrococcus abyssi.
OC   Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX   NCBI_TaxID=29292;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ORSAY;
RA   Heilig R.;
RT   "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
RT   structure and evolution.";
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + TRNA(ASP) = AMP +
CC       PYROPHOSPHATE + L-ASPARTYL-TRNA(ASP).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
```

```
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AJ248288; CAB50576.1; -.
DR   InterPro; IPR002106; AA_tRNA_ligase_II.
DR   InterPro; IPR002318; tRNA-synt_2c.
DR   InterPro; IPR003156; DHHA1.
DR   Pfam; PF02272; DHHA1; 1.
DR   Pfam; PF01411; tRNA-synt_2c; 1.
DR   PRINTS; PR00980; TRNASYNTHALA.
DR   PROSITE; PS00179; AA_TRNA_LIGASE_II_1; FALSE_NEG.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; FALSE_NEG.
KW   Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW   Complete proteome.
SQ   SEQUENCE   914 AA;  104858 MW;  C4B449FE9E03B4E1 CRC64;


  Query Match             22.3%;  Score 52.5;  DB 1;  Length 914;
  Best Local Similarity   35.7%;  Pred. No. 62;
  Matches   15;  Conservative   11;  Mismatches   11;  Indels    5;  Gaps    2;

Qy       6 IDAAIGGSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
           |   : || |   || :::|| ::|   :|| ::|:| ||:
Db     825 IGEIVEGSMDDLREAVERLKKPNRI---VVL--VSKEGYFAI 861


RESULT   9
PCNA_SARCR
ID   PCNA_SARCR     STANDARD;      PRT;   260 AA.
AC   O16852;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PROLIFERATING CELL NUCLEAR ANTIGEN (PCNA) (CYCLIN).
GN   PCNA.
OS   Sarcophaga crassipalpis.
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Oestroidea; Sarcophagidae; Sarcophaga.
OX   NCBI_TaxID=59312;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98382535; PubMed=9714841;
RA   Tammariello S.P., Denlinger D.L.;
RT   "Cloning and sequencing of proliferating cell nuclear antigen (PCNA)
RT   from the flesh fly, Sarcophaga crassipalpis, and its expression in
RT   response to cold shock and heat shock.";
RL   Gene 215:425-429(1998).
CC   -!- FUNCTION: THIS PROTEIN IS AN AUXILIARY PROTEIN OF DNA POLYMERASE
CC       DELTA AND IS INVOLVED IN THE CONTROL OF EUKARYOTIC DNA REPLICATION
CC       BY INCREASING THE POLYMERASE'S PROCESSIBILITY DURING ELONGATION
CC       OF THE LEADING STRAND.
CC   -!- SUBUNIT: HOMOTRIMER. FORMS A COMPLEX WITH ACTIVATOR 1
CC       HETEROPENTAMER IN THE PRESENCE OF ATP (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- SIMILARITY: BELONGS TO THE PCNA FAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AF020427; AAC24238.1; -.
DR   HSSP; P12004; 1AXC.
DR   InterPro; IPR000730; PCNA.
DR   Pfam; PF00705; PCNA; 1.
DR   PRINTS; PR00339; PCNACYCLIN.
DR   ProDom; PD002673; PCNA; 1.
DR   PROSITE; PS01251; PCNA_1; 1.
DR   PROSITE; PS00293; PCNA_2; 1.
KW   DNA-binding; Nuclear protein; DNA replication.
FT   DNA_BIND     61     80       POTENTIAL.
SQ   SEQUENCE   260 AA;  28956 MW;  4645B748EB2996A8 CRC64;


  Query Match             21.7%;  Score 51;  DB 1;  Length 260;
  Best Local Similarity   40.0%;  Pred. No. 25;
  Matches   12;  Conservative    4;  Mismatches   14;  Indels    0;  Gaps    0;

Qy      12 GSADRNAEAFDKMKKDDQIAAAMVLRGMAK 41
           ||  . :: ||| : |  |:  | |  |||
Db      48 GSLTLRSDGFDKFRCDRNISMGMNLGSMAK 77


RESULT  10
TRPD_BRAJA
ID   TRPD_BRAJA     STANDARD;      PRT;   337 AA.
AC   P94326;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   ANTHRANILATE PHOSPHORIBOSYLTRANSFERASE (EC 2.4.2.18).
GN   TRPD.
OS   Bradyrhizobium japonicum.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Bradyrhizobium group; Bradyrhizobium.
OX   NCBI_TaxID=375;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=USDA 110;
RX   MEDLINE=97214643; PubMed=9061023;
RA   Kuykendall L.D., Hunter W.J.;
RT   "The sequence of a symbiotically essential Bradyrhizobium japonicum
RT   operon consisting of trpD, trpC and a moaC-like gene.";
RL   Biochim. Biophys. Acta 1350:277-281(1997).
CC   -!- CATALYTIC ACTIVITY: ANTHRANILATE + PHOSPHORIBOSYLPYROPHOSPHATE =
CC       N-5'-PHOSPHORIBOSYL-ANTHRANILATE + PYROPHOSPHATE.
CC   -!- PATHWAY: SECOND STEP IN BIOSYNTHESIS OF TRYPTOPHAN.
CC   -!- SIMILARITY: BELONGS TO THE ANTHRANILATE PHOSPHORIBOSYLTRANSFERASE
CC       FAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U79771; AAB39009.1; -.
DR   InterPro; IPR003262; Anthr_phosphorbsyltransf.
DR   InterPro; IPR000312; Glycos_transf_3.
DR   InterPro; IPR000053; Thymid_phosphls.
DR   Pfam; PF00591; Glycos_transf_3; 1.
DR   ProDom; PD001864; Glycos_transf_3; 1.
DR   ProDom; PD005916; Thymid_phosphls; 1.
KW   Tryptophan biosynthesis; Transferase; Glycosyltransferase.
SQ   SEQUENCE   337 AA;  34528 MW;  C486203E99E5534E CRC64;


  Query Match             21.7%;  Score 51;  DB 1;  Length 337;
  Best Local Similarity   35.7%;  Pred. No. 33;
  Matches   15;  Conservative    3;  Mismatches   14;  Indels   10;  Gaps    1;

Qy      11 GGSADRNAEAFDKMKKDD----------QIAAAMVLRGMAKD 42
           || || || |   :                |||:|: | |||
Db     266 GGDADANAIALQSVLNGKPSAYRDVALMNAAAALVVAGRAKD 307


RESULT  11
SUCC_ECOLI
```

```
ID   SUCC_ECOLI     STANDARD;      PRT;   388 AA.
AC   P07460;
DT   01-APR-1988 (Rel. 07, Created)
DT   01-APR-1988 (Rel. 07, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   SUCCINYL-COA SYNTHETASE BETA CHAIN (EC 6.2.1.5) (SCS-BETA).
GN   SUCC OR B0728 OR Z0882 OR ECS0753.
OS   Escherichia coli, and
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562, 83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86104124; PubMed=3002435;
RA   Buck D., Spencer M.E., Guest J.R.;
RT   "Primary structure of the succinyl-CoA synthetase of Escherichia
RT   coli.";
RL   Biochemistry 24:6245-6252(1985).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97061202; PubMed=8905232;
RA   Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA   Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA   Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA   Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA   Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA   Yano M., Horiuchi T.;
RT   "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 12.7-28.0 min region on the linkage map.";
RL   DNA Res. 3:137-155(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
RN   [6]
RP   SEQUENCE OF 1-12.
RC   STRAIN=K12 / EMG2;
RX   MEDLINE=97443975; PubMed=9298646;
RA   Link A.J., Robison K., Church G.M.;
RT   "Comparing the predicted and observed properties of proteins encoded
RT   in the genome of Escherichia coli K-12.";
RL   Electrophoresis 18:1259-1313(1997).
RN   [7]
RP   X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX   MEDLINE=94193797; PubMed=8144675;
RA   Wolodko W.T., Fraser M.E., James M.N.G., Bridger W.A.;
RT   "The crystal structure of succinyl-CoA synthetase from Escherichia
RT   coli at 2.5-A resolution.";
RL   J. Biol. Chem. 269:10883-10890(1994).
RN   [8]
RP   X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX   MEDLINE=99141407; PubMed=9917402;
RA   Fraser M.E., James M.N., Bridger W.A., Wolodko W.T.;
RT   "A detailed structural description of Escherichia coli succinyl-CoA
RT   synthetase.";
RL   J. Mol. Biol. 285:1633-1653(1999).
RN   [9]
RP   X-RAY CRYSTALLOGRAPHY (3.3 ANGSTROMS).
RX   MEDLINE=20092606; PubMed=10625475;
RA   Joyce M.A., Fraser M.E., James M.N., Bridger W.A., Wolodko W.T.;
RT   "ADP-binding site of Escherichia coli succinyl-CoA synthetase
RT   revealed by X-ray crystallography.";
RL   Biochemistry 39:17-25(2000).
CC   -!- CATALYTIC ACTIVITY: SUCCINATE + COA + ATP = SUCCINYL-COA + ADP +
CC       ORTHOPHOSPHATE.
CC   -!- ENZYME REGULATION: EXHIBITS TWO INTERESTING PROPERTIES: "SUBSTRATE
CC       SYNERGISM", IN WHICH THE ENZYME IS MOST ACTIVE FOR THE CATALYSIS
CC       OF ITS PARTIAL REACTIONS ONLY WHEN ALL THE SUBSTRATE BINDING SITES
CC       ARE OCCUPIED, AND "CATALYTIC COOPERATIVITY" BETWEEN ALTERNATING
CC       ACTIVE SITES IN THE TETRAMER, WHEREBY THE INTERACTION OF
CC       SUBSTRATES (PARTICULARLY ATP) AT ONE SITE IS NEEDED TO PROMOTE
CC       CATALYSIS AT THE OTHER.
CC   -!- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS.
CC   -!- MISCELLANEOUS: SUCCINYL-COA SYNTHETASE (SCS) OF E.COLI CATALYZES
CC       ITS REACTION VIA THREE STEPS THAT INVOLVE PHOSPHORYL ENZYME AND
CC       ENZYME-BOUND SUCCINYL PHOSPHATE AS INTERMEDIATES.
CC   -!- MISCELLANEOUS: DURING AEROBIC METABOLISM IT FUNCTIONS IN THE
CC       CITRIC ACID CYCLE, COUPLING THE HYDROLYSIS OF SUCCINYL-COA TO THE
CC       SYNTHESIS OF ATP & THUS REPRESENTS AN IMPORTANT SITE OF
CC       SUBSTRATE-LEVEL PHOSPHORYLATION. IT CAN ALSO FUNCTION IN THE
CC       OTHER DIRECTION FOR ANABOLIC PURPOSES, AND THIS MAY BE
CC       PARTICULARLY IMPORTANT FOR PROVIDING SUCCINYL-COA DURING ANAEROBIC
CC       GROWTH WHEN THE OXIDATIVE ROUTE FROM 2-OXOGLUTARATE IS SEVERELY
CC       REPRESSED.
CC   -!- MISCELLANEOUS: THE BETA-SUBUNIT CONTAINS THE ATTACHMENT SITES FOR
CC       SUCCINATE. THE COMPLETE ACTIVE SITE IS PROBABLY LOCATED IN THE
CC       REGION OF ALPHA- BETA CONTACT.
CC   -!- SIMILARITY: TO OTHER BETA SUBUNITS OF SUCCINYL-COA SYNTHETASES.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; J01619; AAA23899.1; -.
DR   EMBL; AE000176; AAC73822.1; -.
DR   EMBL; D90711; BAA35394.1; -.
DR   EMBL; AE005251; AAG55052.1; -.
DR   EMBL; AP002553; BAB34176.1; -.
DR   PIR; A24090; SYECSB.
DR   PDB; 1SCU; 20-APR-95.
DR   PDB; 2SCU; 02-AUG-99.
DR   PDB; 1CQI; 10-JAN-00.
DR   PDB; 1CQJ; 10-JAN-00.
DR   SWISS-2DPAGE; P07460; COLI.
DR   ECO2DBASE; E039.8; 6TH EDITION.
DR   EcoGene; EG10981; sucC.
DR   InterPro; IPR003135; ATP-grasp.
DR   InterPro; IPR000303; CoA_ligase.
DR   Pfam; PF02222; ATP-grasp; 1.
DR   Pfam; PF00549; ligase-CoA; 1.
DR   PROSITE; PS01217; SUCCINYL_COA_LIG_3; 1.
```

```
KW   Ligase; Tricarboxylic acid cycle; 3D-structure; Complete proteome.
SQ   SEQUENCE   388 AA;  41392 MW;  09C429EC97A823CF CRC64;


  Query Match             21.7%;  Score 51;  DB 1;  Length 388;
  Best Local Similarity   34.4%;  Pred. No. 39;
  Matches   11;  Conservative    7;  Mismatches   14;  Indels    0;  Gaps    0;

Qy       4 NPIDAAIGGSADRNAEAFDKMKKDDQIAAAMV 35
           | :|   | : :|  |||  :  ||:: | :|
Db     285 NFLDVGGGATKERVTEAFKIILSDDKVKAVLV 316


RESULT  12
P39_BRUAB
ID   P39_BRUAB      STANDARD;      PRT;   401 AA.
AC   O06875;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   IMMUNOGENIC 39 KDA PROTEIN.
GN   P39.
OS   Brucella abortus.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Brucellaceae; Brucella.
OX   NCBI_TaxID=235;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=544;
RX   MEDLINE=97162316; PubMed=9009303;
RA   Denoel P.A., Vo T.K., Tibor A., Weynants V.E., Trunde J.M., Dubray G.,
RA   Limet J.N., Letesson J.J.;
RT   "Characterization, occurrence, and molecular cloning of a
RT   39-kilodalton Brucella abortus cytoplasmic protein immunodominant in
RT   cattle.";
RL   Infect. Immun. 65:495-502(1997).
CC   -!- FUNCTION: ALLERGENIC IN CATTLE.
CC   -!- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
CC       PROTEIN FAMILY 1.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L35038; AAB50572.1; -.
DR   InterPro; IPR000567; SBP_bac_1.
DR   Pfam; PF01547; SBP_bacterial_1; 1.
DR   PROSITE; PS01037; SBP_BACTERIAL_1; FALSE_NEG.
KW   Allergen; Transport.
SQ   SEQUENCE   401 AA;  42901 MW;  701470A378F338FE CRC64;


  Query Match             21.7%;  Score 51;  DB 1;  Length 401;
  Best Local Similarity   34.1%;  Pred. No. 40;
  Matches   15;  Conservative    7;  Mismatches    8;  Indels   14;  Gaps    2;

Qy       9 AIGGSADRNAEAFDKMKK----------DDQIAAAMVLRGMA 40
           |:|  :|   :|||:| |           |  :|:|||: | |
Db     203 ALG--SDTMKQAFDRMSKLRTYVDDNFSGRDWNLASAMVIEGKA 244


RESULT  13
SYM_ORYSA
ID   SYM_ORYSA      STANDARD;      PRT;   804 AA.
AC   Q9ZTS1;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PROBABLE METHIONYL-TRNA SYNTHETASE (EC 6.1.1.10) (METHIONINE--TRNA
DE   LIGASE) (METRS).
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Deniziak M., Mirande M.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: ATP + L-METHIONINE + TRNA(MET) = AMP +
CC       PYROPHOSPHATE + L-METHIONYL-TRNA(MET).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC   -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF040700; AAC99620.1; -.
DR   InterPro; IPR002300; tRNA-synt_1a.
DR   InterPro; IPR001412; tRNA-synt_I.
DR   InterPro; IPR002304; tRNA-synt_met.
DR   InterPro; IPR002547; tRNA_bind.
DR   Pfam; PF00133; tRNA-synt_1; 1.
DR   Pfam; PF01588; tRNA_bind; 1.
DR   PRINTS; PR01041; TRNASYNTHMET.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW   Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW   tRNA-binding.
FT   DOMAIN        4     11       POLY-PRO.
FT   SITE         28     38       "HIGH" REGION.
FT   SITE        350    354       "KMSKS" REGION.
FT   BINDING     353    353       ATP (BY SIMILARITY).
FT   DOMAIN      648    743       TRNA-BINDING.
SQ   SEQUENCE   804 AA;  89716 MW;  A87E8ABBD419D440 CRC64;


  Query Match             21.7%;  Score 51;  DB 1;  Length 804;
  Best Local Similarity   35.0%;  Pred. No. 84;
  Matches   14;  Conservative    9;  Mismatches   13;  Indels    4;  Gaps    1;

Qy       2 LTNPIDAAIGGSADRNA----EAFDKMKKDDQIAAAMVLR 37
           |   : :|:| |:| ||      | |: |:|  | |:|::
Db     465 LKQGLKSAMGISSDGNAYLQESQFWKLYKEDPAACAVVMK 504


RESULT  14
PSMA_METTE
ID   PSMA_METTE     STANDARD;      PRT;   247 AA.
AC   Q59565;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PROTEASOME ALPHA SUBUNIT (EC 3.4.99.46) (MULTICATALYTIC ENDOPEPTIDASE
DE   COMPLEX ALPHA SUBUNIT).
GN   PSMA.
OS   Methanosarcina thermophila.
OC   Archaea; Euryarchaeota; Methanosarcinales; Methanosarcinaceae;
OC   Methanosarcina.
OX   NCBI_TaxID=2210;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 1825 / TM-1;
RX   MEDLINE=96081920; PubMed=7499378;
RA   Maupin-Furlow J.A., Ferry J.G.;
RT   "A proteasome from the methanogenic archaeon Methanosarcina
RT   thermophila.";
RL   J. Biol. Chem. 270:28617-28622(1995).
```

```
CC   -!- FUNCTION: THE PROTEASOME IS A MULTICATALYTIC PROTEINASE COMPLEX
CC       WHICH IS CHARACTERIZED BY ITS ABILITY TO CLEAVE PEPTIDES WITH
CC       ARG, PHE, TYR, LEU, AND GLU ADJACENT TO THE LEAVING GROUP AT
CC       NEUTRAL OR SLIGHTLY BASIC PH (BY SIMILARITY).
CC   -!- PATHWAY: INVOLVED IN AN ATP/UBIQUITIN-DEPENDENT NON-LYSOSOMAL
CC       PROTEOLYTIC PATHWAY.
CC   -!- SUBUNIT: COMPOSED OF TWO SUBUNITS, ALPHA AND BETA. THE COMPLEX IS
CC       FORMED OF FOUR RINGS. THE TWO OUTER RINGS ARE EACH COMPOSED OF
CC       SEVEN ALPHA SUBUNITS. THE TWO INNER RINGS ARE EACH COMPOSED OF
CC       SEVEN BETA SUBUNITS (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY T1A; ALSO KNOWN AS THE
CC       PROTEASOME A-TYPE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U30483; AAA93166.1; -.
DR   InterPro; IPR001353; Proteasome.
DR   InterPro; IPR000426; Proteasome_A.
DR   Pfam; PF00227; proteasome; 1.
DR   PROSITE; PS00388; PROTEASOME_A; 1.
KW   Proteasome; Hydrolase; Protease.
SQ   SEQUENCE   247 AA;  27155 MW;  B689727743179A5A CRC64;


  Query Match             21.3%;  Score 50;  DB 1;  Length 247;
  Best Local Similarity   30.9%;  Pred. No. 32;
  Matches   17;  Conservative    7;  Mismatches   19;  Indels   12;  Gaps    2;

Qy       3 TNPIDA-------AIGGSADRNAEAFDKMKKDDQIAAAMVLRGM-----AKDGQF 45
           |:|  |       |||   :   | |:   |:|    | :| ||     | :|:|
Db     150 TDPSGALLEYKATAIGAGRNAVVEVFEADYKEDMNIEAAILLGMDALYKAAEGKF 204


RESULT  15
Y024_MYCPN
ID   Y024_MYCPN     STANDARD;      PRT;   362 AA.
AC   P75088;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PROBABLE GTP-BINDING PROTEIN MG024 HOMOLOG (B01_ORF362).
GN   MPN026 OR MP128.
OS   Mycoplasma pneumoniae.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2104;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 29342 / M129;
RX   MEDLINE=97105885; PubMed=8948633;
RA   Himmelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA   Herrmann R.;
RT   "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT   pneumoniae.";
RL   Nucleic Acids Res. 24:4420-4449(1996).
CC   -!- SIMILARITY: BELONGS TO THE YCHF FAMILY OF GTP-BINDING PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000015; AAB95776.1; -.
DR   InterPro; IPR000765; GTP1_OBG.
DR   Pfam; PF01018; GTP1_OBG; 1.
DR   PRINTS; PR00326; GTP1OBG.
KW   GTP-binding; Complete proteome.
FT   NP_BIND       8     15       GTP (POTENTIAL).
FT   NP_BIND      71     75       GTP (POTENTIAL).
SQ   SEQUENCE   362 AA;  40607 MW;  7C79B46C84D5FF83 CRC64;


  Query Match             21.3%;  Score 50;  DB 1;  Length 362;
  Best Local Similarity   35.5%;  Pred. No. 48;
  Matches   11;  Conservative    7;  Mismatches   13;  Indels    0;  Gaps    0;

Qy      12 GSADRNAEAFDKMKKDDQIAAAMVLRGMAKD 42
           |   | ||: ||  |::    | ||:|: ::
Db     144 GKIKRKAESGDKQSKEEYQLLAPVLQGLQQN 174


Search completed: January 10, 2002, 14:09:58
Job time: 327 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 10, 2002, 14:02:26 ; Search time 40.15 Seconds
(without alignments)
643.875 Million cell updates/sec

Title: US-09-445-803-2
Perfect score: 1708
Sequence: 1 KNNDHDNHKGTVKNAVDMAK.....GNGATAKGGDAKSVNGIAKG 349

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101:*
1: /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1708	100.0	349	20	AAW95612	Borrelia burgdorfe
2	1695	99.2	349	21	AAB36281	B. garinii P7-1 pr
3	596	34.9	1328	20	AAY20088	B. burgdorferi ant
4	538.5	31.5	533	20	AAY20112	B. burgdorferi ant
5	437.5	25.6	356	18	AAW22676	Borrelia variable
6	345.5	20.2	156	20	AAY20113	B. burgdorferi ant
7	222	13.0	168	20	AAY20089	B. burgdorferi ant
8	220.5	12.9	738	19	AAW56163	New DNA sequence i
9	214	12.5	681	22	AAB82609	Spider recombinant
10	210.5	12.3	291	22	AAB82608	Spider recombinant
11	208.5	12.2	528	22	AAB82611	Spider recombinant
12	207	12.1	691	22	AAB82610	Spider recombinant
13	203.5	11.9	604	16	AAR99057	Spider dragline va
14	203.5	11.9	606	16	AAR99055	Spider dragline va
15	203.5	11.9	606	20	AAY40101	Polymer of an anal
16	203.5	11.9	606	20	AAY40102	Polymer of an anal
17	197	11.5	651	20	AAY40097	Spider silk protei
18	197	11.5	718	12	AAR14308	N.clavipes draglin
19	197	11.5	718	19	AAW53346	Nephila clavipes s
20	197	11.5	718	21	AAY59070	N. clavipes spider
21	196.5	11.5	831	16	AAR80168	pMISS1 MiSP spider
22	194.5	11.4	606	16	AAR99053	Spider dragline va
23	194.5	11.4	606	20	AAY40100	Polymer of an anal
24	194	11.4	615	20	AAY40099	Spider silk protei
25	193.5	11.3	429	10	AAP90064	Antigenic protein
26	193	11.3	344	17	AAW08097	Vmp7 soluble varia
27	189.5	11.1	646	18	AAW27178	Nephila clavipes s
28	187	10.9	898	18	AAW31853	Mycobacterium tube
29	184.5	10.8	1488	22	AAG91741	C glutamicum prote
30	178	10.4	456	21	AAY67238	Amino acid sequenc
31	177.5	10.4	1822	13	AAR27745	Extracellular fact
32	177	10.4	2411	21	AAB23860	Haemophilus influe
33	176.5	10.3	47	20	AAW95613	Borrelia burgdorfe
34	176.5	10.3	402	8	AAP70709	Plasmodium cynomol
35	172.5	10.1	261	19	AAW79137	FLGA Gly-ala inser
36	171.5	10.0	641	20	AAY28843	Epstein Barr Virus
37	171.5	10.0	641	21	AAY95856	Epstein Barr virus
38	171.5	10.0	641	22	AAB62332	EBV tethering prot
39	170.5	10.0	1018	16	AAW01496	Silk-like protein
40	170.5	10.0	1038	17	AAR95107	Fibronectin cell b
41	170.5	10.0	1038	22	AAB72727	Repetitive protein
42	170.5	10.0	1038	22	AAB63997	FCB-SLP protein fr
43	169.5	9.9	3596	21	AAY87407	Bordetella pertuss
44	169.5	9.9	3647	11	AAR05041	Filamentous haemag
45	169	9.9	1095	21	AAB01835	Haemophilus influe

ALIGNMENTS

RESULT 1
AAW95612
ID AAW95612 standard; Protein; 349 AA.
XX
AC AAW95612;
XX
DT 08-JUN-1999 (first entry)
XX
DE Borrelia burgdorferi surface antigen P39.5 clone 7-1 polypeptide.
XX
KW Lyme disease; surface antigen; P39.5; diagnosis; prevention;
KW vaccine; antisense; therapy; treatment; primer; probe;
KW antibody; DNA.
XX
OS Borrelia burgdorferi.
XX
PN WO9900413-A1.
XX
PD 07-JAN-1999.
XX
PF 29-JUN-1998; 98WO-US13551.
XX
PR 30-JUN-1997; 97US-0051271.
XX
PA (TULA) TULANE EDUCATIONAL FUND.
XX
PI Philipp MT;
XX
DR WPI; 1999-095676/08.
DR N-PSDB; AAX07411.
XX
PT New nucleic acid encoding the P39.5 antigen of Borrelia burgdorferi
PT - and related vectors, transformants, antibodies and polypeptides,
PT for diagnosis, prevention and treatment of Lyme disease

```
XX
PS   Claim 5; Pages 54-55; 89pp; English.
XX
CC   The sequence is that of a Borrelia burgdorferi surface antigen P39.5
CC   clone 7-1 polypeptide. It can be used in the production of P39.5
CC   or fragments of it which may be used to raise antibodies to, and
CC   in the  development of vaccines against Lyme disease. The sequence
CC   can also be used for making primers and probes for diagnosis, also
CC   in DNA vaccines, as antisense therapeutics and for drug screening.
CC   Antibodies can be used as diagnostic (immunoassay) reagents, for
CC   treating Lyme disease, for affinity purification, for drug
CC   screening and to produce anti-idiotype antibodies (used in the
CC   same way as P39.5 to induce an immune response).
XX
SQ   Sequence   349 AA;


  Query Match           100.0%;  Score 1708;  DB 20;  Length 349;
  Best Local Similarity  100.0%;  Pred. No. 1.2e-110;
  Matches  349;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 KNNDHDNHKGTVKNAVDMAKAAEEAASAASAATGNAAIGDVVKNSGAAAKGGEAASVNGI 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 knndhdnhkgtvknavdmakaaeeaasaasaatgnaaigdvvknsgaaakggeaasvngi 60

Qy      61 AKGIKGIVDAAGKADAKEGKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 akgikgivdaagkadakegkldatgaegttnvnagklfvkraaddggdaddagkaaaava 120

Qy     121 ASAATGNAAIGDVVNGDVAKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGAEG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 asaatgnaaigdvvngdvakakggdaasvngiakgikgivdaaekadakegklnaagaeg 180

Qy     181 TTNADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAAD 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ttnadagklfvknagnvggeagdagkaaaavaavsgeqilkaivhaakdggekqgkkaad 240

Qy     241 RTNPIDAAIGGAGDNDAAAAFATMKKDDQIAAAMVLRGMAKDGQFALKDAAAAHEGTVKN 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 rtnpidaaiggagdndaaaafatmkkddqiaaamvlrgmakdgqfalkdaaaahegtvkn 300

Qy     301 AVDIIKAAAEAASAASAATGSAAIGDVVNGNGATAKGGDAKSVNGIAKG 349
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 avdiikaaaeaasaasaatgsaaigdvvngngatakggdaksvngiakg 349


RESULT   2
AAB36281
ID   AAB36281 standard; peptide; 349 AA.
XX
AC   AAB36281;
XX
DT   21-FEB-2001  (first entry)
XX
DE   B. garinii P7-1 protein.
XX
KW   Variable surface antigen; invariable region; VlsE; Lyme disease;
KW   Lyme borreliosis.
XX
OS   Borrelia garinii.
XX
PN   WO200065064-A1.
XX
PD   02-NOV-2000.
XX
PF   25-APR-2000; 2000WO-US11085.
XX
PR   28-APR-1999;   99US-0300971.
XX
PA   (TULA ) TULANE EDUCATIONAL FUND.
XX
PI   Philipp MT,  Liang FT;
XX
DR   WPI; 2000-687350/67.
XX
PT   Novel peptides comprising an invariable 26-amino acid long region
PT   isolated from Borrelia burgdorferi (sensu lato), useful for rapid and
PT   specific diagnosis of lyme disease -
XX
PS   Example 2; Fig 2; 76pp; English.
XX
CC   The present invention describes several peptides comprised of the
CC   invariable regions IR1-IR6 of the B. burgdorferi variable surface antigen
CC   (VlsE) variable domain. These peptides can be used in the specific
CC   diagnosis, treatment and vaccination against B. burgdorferi, which causes
CC   Lyme disease (also known as Lyme borreliosis) in humans, dogs, horses,
CC   cows and other animals.
XX
SQ   Sequence   349 AA;


  Query Match           99.2%;  Score 1695;  DB 21;  Length 349;
  Best Local Similarity  99.4%;  Pred. No. 9.3e-110;
  Matches  347;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 KNNDHDNHKGTVKNAVDMAKAAEEAASAASAATGNAAIGDVVKNSGAAAKGGEAASVNGI 60
           ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 knndhknhkgtvknavdmakaaeeaasaasaatgnaaigdvvknsgaaakggeaasvngi 60

Qy      61 AKGIKGIVDAAGKADAKEGKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 akgikgivdaagkadakegkldatgaegttnvnaqklfvkraaddggdaddagkaaaava 120

Qy     121 ASAATGNAAIGDVVNGDVAKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGAEG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 asaatgnaaigdvvngdvakakggdaasvngiakgikgivdaaekadakegklnaagaeg 180

Qy     181 TTNADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAAD 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ttnadagklfvknagnvggeagdagkaaaavaavsgeqilkaivhaakdggekqgkkaad 240

Qy     241 RTNPIDAAIGGAGDNDAAAAFATMKKDDQIAAAMVLRGMAKDGQFALKDAAAAHEGTVKN 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 atnpidaaiggagdndaaaafatmkkddqiaaamvlrgmakdgqfalkdaaaahegtvkn 300

Qy     301 AVDIIKAAAEAASAASAATGSAAIGDVVNGNGATAKGGDAKSVNGIAKG 349
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 avdiikaaaeaasaasaatgsaaigdvvngngatakggdaksvngiakg 349


RESULT   3
AAY20088
ID   AAY20088 standard; Protein; 1328 AA.
XX
AC   AAY20088;
XX
DT   19-JUL-1999  (first entry)
XX
DE   B. burgdorferi antigenic protein, f24-1.aa.
XX
KW   Antigenic protein; vaccine; Lyme disease; infection; detection.
XX
OS   Borrelia burgdorferi.
XX
PN   WO9859071-A1.
XX
PD   30-DEC-1998.
XX
PF   18-JUN-1998;   98WO-US12718.
XX
PR   03-SEP-1997;   97US-0057483.
PR   20-JUN-1997;   97US-0050359.
PR   22-JUL-1997;   97US-0053344.
PR   22-JUL-1997;   97US-0053377.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
```

PA (MEDI-) MEDIMMUNE INC.
XX
PI Choi GH, Erwin AL, Hanson MS, Lathigra R;
XX
DR WPI; 1999-189980/16.
DR N-PSDB; AAX61785.
XX
PT New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the diagnosis, prevention and treatment of diseases
PT caused by Borrelia, particularly Lyme disease
XX
PS Claim 12; Page 192-193; 275pp; English.
XX
CC This sequence represents a Borrelia burgdorferi (Bb) protein of the
CC invention, which is suitable for use in a vaccine. The Bb polypeptides
CC can be used in vaccines for eliciting protective antibodies to members of
CC the Borrelia genus, particularly for the use against Lyme disease in
CC humans and animals. They can be used for preventing or attenuating an
CC infection caused by a member of the Borrelia genus. The products can also
CC be used for detection of members of the Borrelia genus.
XX
SQ Sequence 1328 AA;

Query Match 34.9%; Score 596; DB 20; Length 1328;
Best Local Similarity 36.5%; Pred. No. 2.6e-33;
Matches 168; Conservative 46; Mismatches 114; Indels 132; Gaps 12;

```
Qy       2 NNDHDNHKGTVKNAVDMA-KAAEEAASAASAATGNAAIGDVVKNSGAAAKGGEAASVNGI 60
           ||: :  :| :|   ::  |      :|  |::|  |||:|| |    ||  : ||| ||
Db     335 nnekekaegaikevselldklvtavktaegassgtdaigevvdn---xakxadkasvtgi 391

Qy      61 AKGIKGIVDAAGKADAKEGKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVA 120
           ||||| ||:||  ::    ||    |    | ||||| |  ||  ||:: | ||| ||:
Db     392 akgikeiveaaxgse----klkvaaaxxxnnkeagklfgkagadangdseaaskaagavs 447

Qy     121 A------------------------------------------------------------ 121
           |
Db     448 avsgeqilsaivkaaaagaadqdgekpgdaknpiaaaigkgnaddgadfgdgmkkddqia 507

Qy     122 ---------------------------------------------SAATGNAAIGDVVN 135
                                                        |::| ||||:||:
Db     508 aaialrgmakdgkfavkkdekgkaegaikgaselldklvkavktaegassgtaaigevvd 567

Qy     136 GDVAKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGAEGTTNADAGKLFVKNAG 195
               ||  |  || ||||||| ||:||  ::    ||  | |:|  |  |||||
Db     568 n---aakaadkdsvtgiakgikeiveaaggse----klkvaaakgennkgagklfgkaga 620

Qy     196 NVGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKD-GGEKQGKKAADRTNPIDAAIGGAGD 254
           |  |::  | ||| ||:|||||||| ||| || :  |:::|||  :  ||| |||   ||
Db     621 nahgdseaaskaagavsavsgeqilsaivkaageaagdqegkkpeeaknpiaaai---gd 677

Qy     255 NDAAAAFAT--MKKDDQIAAAMVLRGMAKDGQFALKDAAAAH--EGTVKNAVDIIKAAAE 310
             | | |    ||||||||||: ||||||||:||:||       || :|   :::    :
Db     678 kdgdaefnqdgmkkddqiaaaialrgmakdgkfavkdggekekaegaikgvselldklvk 737

Qy     311 AA-SAASAATGSAAIGDVVNGNGATAKGGDAKSVNGIAKG 349
           |  :|  |::|:||||:||      ||  |  || |||||
Db     738 avktaegassgtaaigevv---adaakvadkasvtgiakg 774
```

RESULT 4
AAY20112
ID AAY20112 standard; Protein; 533 AA.
XX
AC AAY20112;
XX
DT 19-JUL-1999 (first entry)
XX
DE B. burgdorferi antigenic protein, f49-2.aa.
XX
KW Antigenic protein; vaccine; Lyme disease; infection; detection.
XX
OS Borrelia burgdorferi.
XX
PN WO9859071-A1.
XX
PD 30-DEC-1998.
XX
PF 18-JUN-1998; 98WO-US12718.
XX
PR 03-SEP-1997; 97US-0057483.
PR 20-JUN-1997; 97US-0050359.
PR 22-JUL-1997; 97US-0053344.
PR 22-JUL-1997; 97US-0053377.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMMUNE INC.
XX
PI Choi GH, Erwin AL, Hanson MS, Lathigra R;
XX
DR WPI; 1999-189980/16.
DR N-PSDB; AAX61809.
XX
PT New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the diagnosis, prevention and treatment of diseases
PT caused by Borrelia, particularly Lyme disease
XX
PS Claim 12; Page 201-202; 275pp; English.
XX
CC This sequence represents a Borrelia burgdorferi (Bb) protein of the
CC invention, which is suitable for use in a vaccine. The Bb polypeptides
CC can be used in vaccines for eliciting protective antibodies to members of
CC the Borrelia genus, particularly for the use against Lyme disease in
CC humans and animals. They can be used for preventing or attenuating an
CC infection caused by a member of the Borrelia genus. The products can also
CC be used for detection of members of the Borrelia genus.
XX
SQ Sequence 533 AA;

Query Match 31.5%; Score 538.5; DB 20; Length 533;
Best Local Similarity 38.8%; Pred. No. 9.3e-30;
Matches 157; Conservative 35; Mismatches 94; Indels 119; Gaps 11;

```
Qy      53 EAASVNGIAKGIKGIVDAAGKADAKEGKLDATGAEGTTNVNAGKLFVKRAADDGGDADDA 112
           : ||| |||||||| ||:||| ::    ||    |||  |  ||||| |  | : ||:: |
Db      35 dkasvtgiakgikeiveaaggse----klkvaaaegennekagklfgkagagnagdseaa 90

Qy     113 GKAAAAVAA---------------------------------------------------- 121
             ||| ||:|
Db      91 skaagavsavsgeqilsaivkaageaaqdgekpgeaknpiaaaigkgnedgaefkdemkk 150

Qy     122 ------------------------------------------------SAATGNAAI 130
                                                           |::| |||
Db     151 ddqiaaaialrgmakdgkfavkndekgkaegaikgagelldklvkavktaegassgtaai 210

Qy     131 GDVVNGDVAKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGAEGTTNADAGKLF 190
           |:||  | | ||  | ||| |||||| ||:||    :|: |: ||  ||  |  |||||
Db     211 gevvaddna-akvadkasvkgiakgikeiveaa--ggskklkv-aaakeg--nekagklf 264

Qy     191 VKNAGNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAA-KDGGEKQGKKAADRTNPIDAAI 249
             |      |::  | ||| ||:|||||||| ||| ||    |:::|||  |  ||| |||
Db     265 gkvdaahagdseaaskaagavsavsgeqilsaivkaagaaagdqegkkpgdaknpiaaai 324

Qy     250 G-GAGDNDAAAAFATMKKDDQIAAAMVLRGMAKDGQFALKDAAAAH---EGTVKNAVDII 305
           | |  :| |      ||||||||||: ||||||||:||:|         || :| | :::
Db     325 gkgdaengaefnhdgmkkddqiaaaialrgmakdgkfavksgggekgkaegaikgaaell 384

Qy     306 KAAAEAA-SAASAATGSAAIGDVVNGNGATAKGGDAKSVNGIAKG 349
               :|  :|  |::|: |||:||  |   ||  |  || |||||
Db     385 dklvkavktaegassgtdaigevv-anagaakvadkasvtgiakg 428
```

RESULT 5
AAW22676

```
ID   AAW22676 standard; Protein; 356 AA.
XX
AC   AAW22676;
XX
DT   22-FEB-1998  (first entry)
XX
DE   Borrelia variable major protein (VMP)-like protein VlsE.
XX
KW   Variable major protein-like sequence; VMP-like sequence; vls locus;
KW   VlsE; Lyme disease; relapsing fever; therapy; diagnosis;
KW   vaccine.
XX
OS   Borrelia burgdorferi strain B31-5A3 (ATCC 35210).
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..19
FT                   /label= Sig_peptide
FT                   /note= "lipoprotein signal peptide"
XX
P    WO9731123-A1.
XX
PD   28-AUG-1997.
XX
PF   20-FEB-1997;   97WO-US02952.
XX
PR   21-FEB-1996;   96US-0012028.
XX
PA   (TEXA ) UNIV TEXAS SYSTEM.
XX
PI   Barbour AG,  Hardham JM,  Howell JK,  Norris SJ,  Weinstock GM;
PI   Zhang J;
XX
DR   WPI; 1997-435172/40.
DR   N-PSDB; AAT85042.
XX
PT   Nucleic acid encoding variable major protein-like peptide of
PT   Borrelia  - useful for recombinant production of VMP like protein or
PT   peptide, or for diagnosis of Lyme disease
XX
PS   Claim 2; Page 97-99; 130pp; English.
XX
CC   This protein comprises the surface-exposed lipoprotein variable
CC   major protein (VMP)-like protein VlsE of Borrelia burgdorferi.  Its
CC   sequence was deduced from an isolated vlsE gene (see AAT85042).  An
CC   elaborate genetic system has been identified that promotes
CC   extensive antigenic variation of VlsE.  An infectivity related
CC   28-kb linear plasmid, pBB28La, of B. burgdorferi B31 contains a
C    vmp-like sequence (vls) locus consisting of 15 silent vls cassettes
C    (see AAT85043) and the expressed vlsE gene.  Portions of several of
CC   the 15 silent vls cassette sequences, located approx. 500 bp
CC   upstream of vlsE, recombine into the central vlsE cassette region
CC   during infection, resulting in antigenic variation and hence immune
CC   evasion, long-term survival and pathogenesis in the mammalian host.
CC   Recombinant or native proteins expressed by VMP-like genes, will be
CC   useful for the immunotherapy, immunoprophylaxis and immunodiagnosis
CC   of Lyme disease, relapsing fever and related disorders in humans
CC   and animals.  They can also be used in a Lyme disease vaccine.
XX
SQ   Sequence   356 AA;
```

```
  Query Match             25.6%;  Score 437.5;  DB 18;  Length 356;
  Best Local Similarity   46.9%;  Pred. No. 5.6e-23;
  Matches  127;  Conservative   26;  Mismatches   87;  Indels   31;  Gaps   11;

Qy      79 GKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGDVVNGDV 138
           || |:||: ||         |: |  :  :  |    |   |  |::| ||||:||  |
Db     103 gkpdstgsvgta--------vegaikevselldklvkavktaegassgtaaigevv-ada 153

Qy     139 AKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNA-AGAEGTTNADAGKLFVKNAGNV 197
             ||  | ||| ||||||| ||:|| ::    || | | |:|  |  ||||| |
Db     154 daakvadkasvkgiakgikeiveaaggse----klkavaaakgennkgagklfgkagaaa 209
```

```
Qy     198 GGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAADRTNPIDAAIGGAGDNDA 257
           |::  | ||| ||:||||||| ||| || |  |: |||  :  ||| |||   || |
Db     210 hgdseaaskaagavsavsgeqilsaivtaa-daaeqdgkkpeeaknpiaaai---gdkdg 265

Qy     258 AAAFA--TMKKDDQIAAAMVLRGMAKDGQFALKDAAAAHEGTVKNAVDIIKAAAEAA--S 313
             | |    |||||||||: |||||||:||:||      |  : |   || |||:|
Db     266 gaefgqdemkkddqiaaaialrgmakdgkfavkd------gekekaegaikgaaesavrk 319

Qy     314 AASAATGSAAIGDVVNGNGATAKGGDAKSVN 344
             | ||   ||| |: :|    |   |: :
Db     320 vlgaitg--ligdavs-sglrkvgdsvkaas 347
```

```
RESULT   6
AAY20113
ID   AAY20113 standard; Protein; 156 AA.
XX
AC   AAY20113;
XX
DT   19-JUL-1999  (first entry)
XX
DE   B. burgdorferi antigenic protein, t49-2.aa.
XX
KW   Antigenic protein; vaccine; Lyme disease; infection; detection.
XX
OS   Borrelia burgdorferi.
XX
PN   WO9859071-A1.
XX
PD   30-DEC-1998.
XX
PF   18-JUN-1998;   98WO-US12718.
XX
PR   03-SEP-1997;   97US-0057483.
PR   20-JUN-1997;   97US-0050359.
PR   22-JUL-1997;   97US-0053344.
PR   22-JUL-1997;   97US-0053377.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Choi GH,  Erwin AL,  Hanson MS,  Lathigra R;
XX
DR   WPI; 1999-189980/16.
DR   N-PSDB; AAX61810.
XX
PT   New isolated Borrelia burgdorferi nucleic acids  - used to develop
PT   products for the diagnosis, prevention and treatment of diseases
PT   caused by Borrelia, particularly Lyme disease
XX
PS   Claim 12; Page 202; 275pp; English.
XX
CC   This sequence represents a Borrelia burgdorferi (Bb) protein of the
CC   invention, which is suitable for use in a vaccine. The Bb polypeptides
CC   can be used in vaccines for eliciting protective antibodies to members of
CC   the Borrelia genus, particularly for the use against Lyme disease in
CC   humans and animals. They can be used for preventing or attenuating an
CC   infection caused by a member of the Borrelia genus. The products can also
CC   be used for detection of members of the Borrelia genus.
XX
SQ   Sequence   156 AA;
```

```
  Query Match             20.2%;  Score 345.5;  DB 20;  Length 156;
  Best Local Similarity   53.4%;  Pred. No. 5.2e-17;
  Matches   86;  Conservative   17;  Mismatches   51;  Indels    7;  Gaps    4;

Qy     140 KAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGAEGTTNADAGKLFVKNAGNVGG 199
           |::  | ||| ||||||| ||:||  ::    ||  | |||  |  ||||| |      |
Db       2 ksqvadkasvtgiakgikeiveaaggse----klkvaaaegennekagklfgkagagnag 57

Qy     200 EAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAADRTNPIDAAIGGAGDNDAAA 259
           ::  | ||| ||:||||||| ||| || :  :  |:|  :  ||| ||| | |: | |
```

```
Db     58 dseaaskaagavsavsgeqilsaivkaageaaq-dgekpgeaknpiaaai-gkgnedgae 115

Qy    260 AFATMKKDDQIAAAMVLRGMAKDGQFALK-DAAAAHEGTVK 299
Db    116 fkdemkkddqiaaaialrgmakdgkfavkndekgkaegaik 156


RESULT   7
AAY20089
ID   AAY20089 standard; Protein; 168 AA.
XX
AC   AAY20089;
XX
DT   19-JUL-1999  (first entry)
XX
DE   B. burgdorferi antigenic protein, t24-1.aa.
XX
KW   Antigenic protein; vaccine; Lyme disease; infection; detection.
XX
OS   Borrelia burgdorferi.
XX
PN   WO9859071-A1.
XX
PD   30-DEC-1998.
XX
PF   18-JUN-1998;   98WO-US12718.
XX
PR   03-SEP-1997;   97US-0057483.
PR   20-JUN-1997;   97US-0050359.
PR   22-JUL-1997;   97US-0053344.
PR   22-JUL-1997;   97US-0053377.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Choi GH,  Erwin AL,  Hanson MS,  Lathigra R;
XX
DR   WPI; 1999-189980/16.
DR   N-PSDB; AAX61786.
XX
PT   New isolated Borrelia burgdorferi nucleic acids  - used to develop
PT   products for the diagnosis, prevention and treatment of diseases
PT   caused by Borrelia, particularly Lyme disease
XX
PS   Claim 12; Page 193; 275pp; English.
XX
CC   This sequence represents a Borrelia burgdorferi (Bb) protein of the
CC   invention, which is suitable for use in a vaccine. The Bb polypeptides
CC   can be used in vaccines for eliciting protective antibodies to members of
CC   the Borrelia genus, particularly for the use against Lyme disease in
CC   humans and animals. They can be used for preventing or attenuating an
CC   infection caused by a member of the Borrelia genus. The products can also
CC   be used for detection of members of the Borrelia genus.
XX
SQ   Sequence   168 AA;


  Query Match             13.0%;  Score 222;  DB 20;  Length 168;
  Best Local Similarity   45.5%;  Pred. No. 1.9e-08;
  Matches   55;  Conservative   17;  Mismatches   43;  Indels    6;  Gaps    3;

Qy    232 EKQGKKAADRTNPIDAAIGGAGDNDAAAAFATMKKDDQIAAAMVLRGMAKDGQFALK--D 289
Db      4 eqdgekpedaknpiaaaigkgngdgaefdqdemkkddqiaaaialrgmakdgkfavkgnn 63

Qy    290 AAAAHEGTVKNAVDII-KAAAEAASAASAATGSAAIGDVVNGNGATAKGGDAKSVNGIAK 348
Db     64 ekekaegaikevselldklvtavktaegassgtdaigevvdn---xakxadkasvtgiak 120

Qy    349 G 349
Db    121 g 121


RESULT   8
AAW56163
ID   AAW56163 standard; Protein; 738 AA.
XX
AC   AAW56163;
XX
DT   28-JUL-1998  (first entry)
XX
DE   New DNA sequence isolated from Pinctada fucata.
XX
KW   Pinctada fucata; protein; cosmetic.
XX
OS   Pinctada fucata.
XX
PN   JP10080285-A.
XX
PD   31-MAR-1998.
XX
PF   28-MAY-1997;   97JP-0138461.
XX
PR   15-JUL-1996;   96JP-0184459.
XX
PA   (MIKI-) MIKIMOTO SEIYAKU KK.
XX
DR   WPI; 1998-254410/23.
DR   N-PSDB; AAV22683.
XX
PT   New cDNA and e.g. vector, host cell and polypeptide - used to
PT   produce polypeptide in high yields, which is used in cosmetics
XX
PS   Claim 9; Pages 9-11; 15pp; Japanese.
XX
CC   The present sequence represents protein encoded by a new DNA sequence
CC   isolated from Pinctada fucata. The protein be used as an ingredient
CC   in cosmetics.
XX
SQ   Sequence   738 AA;


  Query Match             12.9%;  Score 220.5;  DB 19;  Length 738;
  Best Local Similarity   29.6%;  Pred. No. 1.2e-07;
  Matches  113;  Conservative   29;  Mismatches  167;  Indels   73;  Gaps   14;

Qy     19 AKAAEEAASAASAATGNAAIGDVVKNSGAAAKGGEAASVNGIAKGIKGIVDAAGKADAKE 78
Db    324 aaaaaaaaaaaggggggggggggggagaaaaaaaaaasasasrqmsgirdalg--dikd 381

Qy     79 GKLDATGAEGTTNVNAGKL-FVKRAADDGGDA--DDAG------KAAAAVAASAATGNAA 129
Db    382 -llrsngasakasakasavastksqiddlkdvlkdlaglkssasasasasasasasagggg 440

Qy    130 IGDVVNGDVAKAKGGDAASVNGI-AKGIKGIVDAAEKADAKEGKLNAAGAEGTTNADAGK 188
Db    441 gggnggggnggggggagalaaalaaagagggglgggggggalaaalaaagaggggfgglgg 500

Qy    189 LFVKNAGNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAADRTNPIDAA 248
Db    501 l-----gglgggsaaaaaaaaaaasggggralrralrrqmrgg---gsaaaaaaaaaaaa 552

Qy    249 IGG----------------------AGDNDAAAAFATMKKDDQIAAAMVL---------RG 278
Db    553 gggwgggmgggfgvglgggfgggfgggssaaaaaaa-------aaaagfggggrrgrgrg 605

Qy    279 MAKDGQ----FALKDAAAAHEGTVKNAVDIIKAAAEAASAASAATGSAAIGDVV------ 328
Db    606 rggdgdgngasavaaaaaaaaaaggsaadv----aaaaaaaaamygdgadgpdfdngfgg 661

Qy    329 -NGNGATAKGGDAKSVNGIAKG 349
Db    662 gngngggggsggggsggggsggg 683
```

```
RESULT   9
AAB82609
ID   AAB82609 standard; Protein; 681 AA.
XX
AC   AAB82609;
XX
DT   02-OCT-2001  (first entry)
XX
DE   Spider recombinant silk protein pQE((SP1)4/(SP2)1)4.
XX
KW   Spider; orb-weaver; silk protein; pQE((SP1)4/(SP2)1)4;
KW   structural protein; purification; fibre; spinning.
XX
OS   Nephila clavipes.
XX
PN   WO200153333-A1.
XX
PD   26-JUL-2001.
XX
PF   01-NOV-2000; 2000WO-US30086.
XX
PR   20-JAN-2000; 2000US-0490291.
XX
PA   (MELL/) MELLO C M.
PA   (ARCI/) ARCIDIACONO S.
PA   (BUTL/) BUTLER M M.
PA   (USSA ) US SEC OF ARMY.
XX
PI   Mello CM,  Arcidiacono S,  Butler MM;
XX
DR   WPI; 2001-483136/52.
DR   N-PSDB; AAH26302.
XX
PT   Recovering structural polypeptides in a biological sample, useful for
PT   purifying and spinning spider silks and other structural proteins,
PT   comprises treating the sample containing the polypeptides with an acid
PT   -
XX
PS   Claim 2; Page 34-37; 49pp; English.
XX
CC   The present sequence is that of orb-weaver spider (Nephila clavipes)
CC   recombinant silk protein pQE((SP1)4/(SP2)1)4.  The invention
CC   provides methods for purifying and spinning spider silks and other
CC   structural proteins.  Organic acids are used to lyse recombinant
CC   cells or other biological samples (such as non-recombinantly
CC   derived cells), and enrich the purity and yields of structural
CC   proteins by hydrolysing many of the macromolecules while leaving
CC   the structural proteins intact.  In the case of silk proteins, the
CC   resulting lysate is further purified by ion-exchange or affinity
CC   chromatography and processed into an aqueous-based mixture for
CC   fibre spinning.  In the present case, the pQE((SP1)4/(SP2)1)4 gene
CC   was cloned into vector pQE-9 for recombinant expression in
CC   Escherichia coli, and recombinant silk protein was obtained in 97%
CC   purity using propionic acid and anion-exchange chromatography on
CC   QAE-Sephadex A50.  Products obtained using the methods of the
CC   invention can be used in the construction of many materials
CC   including films, fibres, woven articles, sutures, ballistic
CC   protection, parachutes and parachute cords.  The new method has
CC   the following advantages over prior art: it involves fewer steps,
CC   requires less time and smaller volumes of reagents, results in
CC   better recovery of protein at higher purity (70-99%), is easy to
CC   scale up, and the fibres are spun in an environmentally benign
CC   solution reducing hazardous waste accumulation and cost.
XX
SQ   Sequence   681 AA;

  Query Match             12.5%;  Score 214;  DB 22;  Length 681;
  Best Local Similarity   26.6%;  Pred. No. 3e-07;
  Matches   93;  Conservative   41;  Mismatches  151;  Indels   64;  Gaps   15;

Qy       1 KNNDHDNHKGTVKNAVD--MAKAAEEAASAASAATGNAAIGDV--VKNSGAAAKGGEAAS 56
           : : | :| |:: :       : |  ||:||:|| | |  |    : : | : :||
Db       2 rgshhhhhhgsmasgrgglggqgagaaaaaaaaaaggagqggygglgsqgtsgrgg---- 57

Qy      57 VNGIAKGIKGIVDAAGKADAKEGKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAA 116
           : |   |      ||    | :|    |::||:             |     |  |
Db      58 lgggagaaaaaaaaaaggagqggygglgsqgtsg-------------rgglgggagaa 103

Qy     117 AAVAASAATGNAAIGDV--VNGDVAKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLN 174
           || ||:|| | |  |    :       :||    : |  |      ||    | :|
Db     104 aaaaaaaaaggagqggygglgsqgtsgrgg----lgggagaaaaaaaaaaggagqggyg 159

Qy     175 AAGAEGTTNADA---GKLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGG 231
             |::||:       |:      | :||:   || |||| || :|          |  ||
Db     160 glgsqgtsgpggygpgqg-tsgrgglgggg--agaaaaaaaaaag---------gagggg 207

Qy     232 EKQGKKAADRTNPIDAAIGGAGDNDAAAAFATMKKDDQIAAAMVLRGMAKDGQFALKDAA 291
             |   : |    :  || |   | || |        |||    |  : |   |
Db     208 --ygglgsqgt----sgrgglgggagaaaa---------aaaaaaggagqggygglgsqg 253

Qy     292 AAHEGTVKNAVDIIKAAAEAASAASAATGSAAIGDV--VNGNGATAKGG 338
           :  | :       : |  ||:||:|| | |  |    :  | : :||
Db     254 tsgrgglgg-----qgagaaaaaaaaaaggagqggygglgsqgtsgrgg 297


RESULT  10
AAB82608
ID   AAB82608 standard; Protein; 291 AA.
XX
AC   AAB82608;
XX
DT   02-OCT-2001  (first entry)
XX
DE   Spider recombinant silk protein pQE(sp1)7.
XX
KW   Spider; orb-weaver; silk protein; pQE(sp1)7; structural protein;
KW   purification; fibre; spinning.
XX
OS   Nephila clavipes.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 232
FT                   /note= "encoded by GGY"
XX
PN   WO200153333-A1.
XX
PD   26-JUL-2001.
XX
PF   01-NOV-2000; 2000WO-US30086.
XX
PR   20-JAN-2000; 2000US-0490291.
XX
PA   (MELL/) MELLO C M.
PA   (ARCI/) ARCIDIACONO S.
PA   (BUTL/) BUTLER M M.
PA   (USSA ) US SEC OF ARMY.
XX
PI   Mello CM,  Arcidiacono S,  Butler MM;
XX
DR   WPI; 2001-483136/52.
DR   N-PSDB; AAH26301.
XX
PT   Recovering structural polypeptides in a biological sample, useful for
PT   purifying and spinning spider silks and other structural proteins,
PT   comprises treating the sample containing the polypeptides with an acid
PT   -
XX
PS   Claim 2; Page 33-34; 49pp; English.
XX
CC   The present sequence is that of orb-weaver spider (Nephila clavipes)
CC   recombinant silk protein, pQE(sp1)7.  The invention provides
CC   methods for purifying and spinning spider silks and other
CC   structural proteins.  Organic acids are used to lyse recombinant
CC   cells or other biological samples (such as non-recombinantly
CC   derived cells), and enrich the purity and yields of structural
```

```
CC   proteins by hydrolysing many of the macromolecules while leaving
CC   the structural proteins intact.  In the case of silk proteins, the
CC   resulting lysate is further purified by ion-exchange or affinity
CC   chromatography and processed into an aqueous-based mixture for
CC   fibre spinning.  In the present case, the pQE(sp1)7 gene was cloned
CC   into vector pQE-9 for recombinant expression in Escherichia coli,
CC   and recombinant silk protein was purified using formic acid and
CC   ion-exchange chromatography.  Products obtained using the methods
CC   of the invention can be used in the construction of many materials
CC   including films, fibres, woven articles, sutures, ballistic
CC   protection, parachutes and parachute cords.  The new method has
CC   the following advantages over prior art: it involves fewer steps,
CC   requires less time and smaller volumes of reagents, results in
CC   better recovery of protein at higher purity (70-99%), is easy to
CC   scale up, and the fibres are spun in an environmentally benign
CC   solution reducing hazardous waste accumulation and cost.
XX
SQ   Sequence   291 AA;


 [redacted]ry Match            12.3%;  Score 210.5;  DB 22;  Length 291;
  [redacted]t Local Similarity   26.7%;  Pred. No. 2.1e-07;
  Matches   92;  Conservative   38;  Mismatches  145;  Indels   69;  Gaps   14;

Qy       1 KNNDHDNHKGTVKNAVD--MAKAAEEAASAASAATGNAAIGDV--VKNSGAAAKGGEAAS 56
           : : | :| |:: :       : |  ||:||:|| | |  |    : : | : :||
Db       2 rgshhhhhhgsmasgrgglggqgagaaaaaaaaaggagqggygglgsqgtsgrgg---- 57

Qy      57 VNGIAKGIKGIVDAAGKADAKEGKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAA 116
           : |   |      ||    | :|     |::||:              |     |  |
Db      58 lgqqgagaaaaaaaaaaggagqgggygglgsqgtsg-------------rgglgqgqaga 103

Qy     117 AAVAASAATGNAAIGDV--VNGDVAKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLN 174
           || ||:|| | |  |    :       :||    : |   |      ||    | :|
Db     104 aaaaaaaaaggagqggygglgsqgtsgrgg----lgqqgagaaaaaaaaaaggagqgggyg 159

Qy     175 AAGAEGTTNADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQ 234
             |::||:            | :||:   || |||| || :|          |  ||
Db     160 glgsqgts----------grgglgqgg--agaaaaaaaaaag---------gagqgg--y 196

Qy     235 GKKAADRTNPIDAAIGGAGDNDAAAAFATMKKDDQIAAAMVLRGMAKDGQFALKDAAAAH 294
           |   :  |    :  || |   | || |        |||    |  : |   |     :
Db     197 gglgsqgt----sgrgglgqggagaaaa--------aaaaaaggagqggygglgsqgtsg 244

Qy     295 EGTVKNAVDIIKAAAEAASAASAATGSAAIGDVVNGNGATAKGG 338
           | :       : |  ||:||:|| | |  |    | |:    |
Db     245 rgglgg-----qgagaaaaaaaaaaggagqgg-ygglgsqgtsg 282


 [redacted]T  11
 [redacted]611
ID   AAB82611 standard; Protein; 528 AA.
XX
AC   AAB82611;
XX
DT   02-OCT-2001  (first entry)
XX
DE   Spider recombinant silk protein pETNcDS.
XX
KW   Spider; orb-weaver; silk protein; pETNcDS; structural protein;
KW   purification; fibre; spinning.
XX
OS   Nephila clavipes.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 417
FT                   /note= "encoded by TAT"
FT   Misc-difference 427
FT                   /note= "encoded by CGA"
FT   Misc-difference 522
FT                   /note= "encoded by GAG"
XX
PN   WO200153333-A1.
XX
PD   26-JUL-2001.
XX
PF   01-NOV-2000; 2000WO-US30086.
XX
PR   20-JAN-2000; 2000US-0490291.
XX
PA   (MELL/) MELLO C M.
PA   (ARCI/) ARCIDIACONO S.
PA   (BUTL/) BUTLER M M.
PA   (USSA ) US SEC OF ARMY.
XX
PI   Mello CM,  Arcidiacono S,  Butler MM;
XX
DR   WPI; 2001-483136/52.
DR   N-PSDB; AAH26304.
XX
PT   Recovering structural polypeptides in a biological sample, useful for
PT   purifying and spinning spider silks and other structural proteins,
PT   comprises treating the sample containing the polypeptides with an acid
PT   -
XX
PS   Claim 2; Page 41-42; 49pp; English.
XX
CC   The present sequence is that of the orb-weaver spider (Nephila
CC   clavipes) recombinant silk protein pETNcDS.  The invention
CC   provides methods for purifying and spinning spider silks and other
CC   structural proteins.  Organic acids are used to lyse recombinant
CC   cells or other biological samples (such as non-recombinantly
CC   derived cells), and enrich the purity and yields of structural
CC   proteins by hydrolysing many of the macromolecules while leaving
CC   the structural proteins intact.  In the case of silk proteins, the
CC   resulting lysate is further purified by ion-exchange or affinity
CC   chromatography and processed into an aqueous-based mixture for
CC   fibre spinning.  In the present case, the pETNcDS gene was cloned
CC   into vector pET24 for recombinant expression in Escherichia coli,
CC   and recombinant silk protein was obtained using formic acid
CC   containing denaturant (3 M guanidine-HCl) and by affinity
CC   chromatography on nickel-NTA agarose.  Products obtained using the
CC   methods of the invention can be used in the construction of many
CC   materials including films, fibres, woven articles, sutures,
CC   ballistic protection, parachutes and parachute cords.  The new
CC   method has the following advantages over prior art: it involves
CC   fewer steps, requires less time and smaller volumes of reagents,
CC   results in better recovery of protein at higher purity (70-99%),
CC   is easy to scale up, and the fibres are spun in an environmentally
CC   benign solution reducing hazardous waste accumulation and cost.
XX
SQ   Sequence   528 AA;


  Query Match            12.2%;  Score 208.5;  DB 22;  Length 528;
  Best Local Similarity  27.8%;  Pred. No. 5.4e-07;
  Matches   93;  Conservative   23;  Mismatches  174;  Indels   45;  Gaps   11;

Qy      24 EAASAASAATGNAAIGDVVKNSGAAAKGGEAASVNGIAKGIKGIVDAAGKADAKEGKLDA 83
           : | ||:|| | |  |      |    | : |   |:     |   ||    | :|
Db     148 qgagaaaaaaggagqg------gygglgsqgagrgglgggagaaaaaaaggagqgggygg 201

Qy      84 TGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGDVVNGDVAKAKG 143
             | :|      | |     :   |     |: | | ||:   |   :|    |  | | |
Db     202 lgggqagqggyggl----gsqgagrgglgggagaaaaaggagqggglgggagaaaaaag 257

Qy     144 G------DAASVNGIAKGIKGIVDAAEKA-DAKEGKLNAAGAEGTTNADAGKLFVKNAGN 196
           |           |  :| :|   ||  |  | :|     | :|      | |  : ||
Db     258 gagqggygglgsqgagrgggagaaaaaaggagqg---gyggqgagqggygglgsqgagr 314

Qy     197 VGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKD--GGEKQGKKAADRTNPIDAAIGGAGD 254
             ||  |    |||| ||  |         | :   ||:  |  ||         :|| |
Db     315 -gglgggagaaaaaaaagg---------agggglgggagaaaaaaggagqggglgggga 364

Qy     255 NDAAAAFATMKKDDQIAAAMVLRGMAKDGQFALKDAAAAHEGTVKNAVDIIKAAAEAASA 314
             | | |       |||    |: : |   |   | .|    |     ||| |  |
```

```
Db    365 gqgagaaaa-------aaaaaaggvrqggygglgsqgagrggggaga-----aaaaagga 412

Qy    315 ASAATGSAAIGDVVNGNGATAKGGDAKSVNGIAKG 349
          ||    |  |   |   :|  | :  |  :|
Db    413 ggggtgglg-gggvgagglgggagaaaavgaggg 446


RESULT  12
AAB82610
ID   AAB82610 standard; Protein; 691 AA.
XX
AC   AAB82610;
XX
DT   02-OCT-2001  (first entry)
XX
DE   Spider recombinant silk protein pET((SP1)4/(SP2)1)4.
XX
KW   Spider; orb-weaver; silk protein; pET((SP1)4/(SP2)1)4;
KW   structural protein; purification; fibre; spinning.
XX
OS   Nephila clavipes.
XX
PN   WO200153333-A1.
XX
PD   26-JUL-2001.
XX
PF   01-NOV-2000; 2000WO-US30086.
XX
PR   20-JAN-2000; 2000US-0490291.
XX
PA   (MELL/) MELLO C M.
PA   (ARCI/) ARCIDIACONO S.
PA   (BUTL/) BUTLER M M.
PA   (USSA ) US SEC OF ARMY.
XX
PI   Mello CM,  Arcidiacono S,  Butler MM;
XX
DR   WPI; 2001-483136/52.
DR   N-PSDB; AAH26303.
XX
PT   Recovering structural polypeptides in a biological sample, useful for
PT   purifying and spinning spider silks and other structural proteins,
PT   comprises treating the sample containing the polypeptides with an acid
PT   -
XX
PS   Claim 2; Page 38-40; 49pp; English.
XX
CC   The present sequence is that of orb-weaver spider (Nephila clavipes)
CC   recombinant silk protein pET((SP1)4/(SP2)1)4.  The invention
CC   provides methods for purifying and spinning spider silks and other
CC   structural proteins.  Organic acids are used to lyse recombinant
CC   cells or other biological samples (such as non-recombinantly
CC   derived cells), and enrich the purity and yields of structural
CC   proteins by hydrolysing many of the macromolecules while leaving
CC   the structural proteins intact.  In the case of silk proteins, the
CC   resulting lysate is further purified by ion-exchange or affinity
CC   chromatography and processed into an aqueous-based mixture for
CC   fibre spinning.  In the present case, the pET((SP1)4/(SP2)1)4 gene
CC   was cloned into vector pET24 for recombinant expression in
CC   Escherichia coli, and recombinant silk protein was obtained in
CC   75-75% purity using propionic acid and anion-exchange chromatography
CC   on QAE-Sephadex A50.  When denaturant (3 M guanidine-HCl) was added
CC   to the propionic acid, the recombinant protein was obtained in 80%
CC   purity.  Products obtained using the methods of the invention can
CC   be used in the construction of many materials including films,
CC   fibres, woven articles, sutures, ballistic protection, parachutes
CC   and parachute cords.  The new method has the following advantages
CC   over prior art: it involves fewer steps, requires less time and
CC   smaller volumes of reagents, results in better recovery of protein
CC   at higher purity (70-99%), is easy to scale up, and the fibres are
CC   spun in an environmentally benign solution reducing hazardous waste
CC   accumulation and cost.
XX

SQ   Sequence   691 AA;


  Query Match             12.1%;  Score 207;  DB 22;  Length 691;
  Best Local Similarity   27.4%;  Pred. No. 9.2e-07;
  Matches   90;  Conservative   35;  Mismatches  141;  Indels   62;  Gaps   14;

Qy     20 KAAEEAASAASAATGNAAIGDV--VKNSGAAAKGGEAASVNGIAKGIKGIVDAAGKADAK 77
          : |  ||:||:|| | |  |    : : | : :||    : |   |      ||    |
Db     25 qgagaaaaaaaaaaggagqggyygglgsqgtsgrgq----lgggagaaaaaaaaaaggag 80

Qy     78 EGKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGDV--VN 135
          :|     |::||:                |     |  ||| ||:|| | |  |    :
Db     81 qggygglgsqgtsg--------------rgglgqqgagaaaaaaaaaaggagqgggyyglg 126

Qy    136 GDVAKAKGGDAASVNGIAKGIKGIVDAAEKADAKGKLNAAGAEGTTNADA---GKLFVK 192
             :||    : |   |      ||    | :|     |::||:       |:
Db    127 sqgtsgrgg----lgggagaaaaaaaaaaggagqggyygglgsqgtsgpggygpgqq-ts 181

Qy    193 NAGNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAADRTNPIDAAIGGA 252
            | :||:   || |||| || :|          |  ||   |   :  |    :  ||
Db    182 grgglgggg--agaaaaaaaaaag---------gagqgg--ygglgsqgt----sgrggl 224

Qy    253 GDNDAAAAFATMKKDDQIAAAMVLRGMAKDGQFALKDAAAAHEGTVKNAVDIIKAAAEAA 312
          |   | || |         |||    |  : |   |     :  | :       : |  ||
Db    225 ggggagaaaa--------aaaaaaggagqggyygglgsqgtsgrgglgg-----qgagaaa 271

Qy    313 SAASAATGSAAIGDV--VNGNGATAKGG 338
          :||:|| | |  |    :   | : :||
Db    272 aaaaaaaggagqggyygglgsqgtsgrgg 299


RESULT  13
AAR99057
ID   AAR99057 standard; Protein; 604 AA.
XX
AC   AAR99057;
XX
DT   17-JAN-1997  (first entry)
XX
DE   Spider dragline variant, DP-1B.16 polymer.
XX
KW   Spider; dragline protein; variant;  monomer; polymer;
KW   fibre forming region; Spidroin 1; Nephila clavipes; DP1; mimic;
KW   DP-1A analogue; fibre; high tensile strength; elasticity; clothing;
KW   rope; surgical suture; implant; reinforcement; film; coating.
XX
OS   Synthetic.
XX
PN   WO9429450-A2.
XX
PD   22-DEC-1994.
XX
PF   15-JUN-1994;   94WO-US06689.
XX
PR   15-JUN-1993;   93US-0077600.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Fahnestock SR;
XX
DR   WPI; 1995-036479/05.
XX
PT   New synthetic variants of spider dragline protein - for making
PT   fibres useful as clothing, surgical silk, plastic reinforcement
PT   etc., also related DNA, vectors and transformed cells
XX
PS   Disclosure; Page 106-108; 168pp; English.
XX
CC   This sequence represents a synthetic spider dragline variant polymer,
CC   DP-1B.16.  The sequence of the DP-1B.16 monomer is given in AAR99056.
CC   The polypeptide monomer is a variant based on a consensus sequence
CC   derived from the fibre forming regions of spider dragline protein,
```

```
CC   esp. the natural protein 1 (Spidroin 1) from Nephila clavipes.  The
CC   DNA sequence encoding the monomer may be used in the recombinant
CC   production of the variant protein in a recombinant host, e.g. E. coli
CC   or Bacillus subtilis.  Synthetic analogues of DP1 were designed to mimic
CC   the repeating consensus sequence of the natural protein and the pattern
CC   of variation among individual repeats.  This monomer exhibits all of the
CC   regularities of (1)-(5) below.  In addition, it exhibits a regularity of
CC   the natural sequence which is not shared by DP-1A, namely that a repeat
CC   in which both GYG and GRG are deleted is generally preceded by a repeat
CC   lacking the entire poly-alanine repeat, with one intervening repeat.
CC   The sequence of DP-1B matches the natural sequence more closely over
CC   a more extended segment than does DP-1A.  The individual repeats differ
CC   from the consensus sequence given in AAW06201 according to the pattern:
CC          (1) the poly-alanine sequence varies in length from 0-7
CC          residues; (2) when the entire poly-alanine sequence is deleted,
CC   so also is the surrounding sequence encompassing AGRGGLGGQGAGAnGG;
CC          (3) aside from the poly-alanine sequence, deletions usually
CC   encompass integral multiples of three consecutive residues;
CC          (4) deletion of GYG is generally accompanied by deletion of GRG
CC   in the same sequence; and
CC          (5) a repeat in which the entire poly-alanine sequence is
CC   deleted is generally preceded by a repeat containing six alanine
CC   residues.
CC   The proteins may be used to produce fibres of high tensile strength and
CC   elasticity, suitable for clothing, rope, surgical sutures, biomaterials
CC   for implants, plastic reinforcements, films, coatings, etc.
XX
SQ   Sequence   604 AA;


  Query Match             11.9%;  Score 203.5;  DB 16;  Length 604;
  Best Local Similarity   26.6%;  Pred. No. 1.4e-06;
  Matches  102;  Conservative   35;  Mismatches  162;  Indels   85;  Gaps   13;

Qy      19 AKAAEEAASAASAATGNA------AIGDVVKNSGAAAKGGEAASVNGIAKGIKGIVDAAG 72
           :: | : | ||:|| | |       :|      |    |  ||:  |   |  |:   ||
Db     247 sqgagggagaaaaaaggagggggygglgsqgagrgggggagaaaaaaggagggygglgsgag 306

Qy      73 KADAKEGKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGD 132
             :|     |::|     ||:          ||     |   |  ||:|| | |  |
Db     307 -----qggygglgsqg-----agr---------gg----lggqgagaaaaaaaggagqgg 343

Qy     133 VVNGDVAKAKGGDAASVNGIAKGIKGIVD--------------AAEKADAKEGKLNAAGA 178
           : :    :  |  ||:  |  :|  | :               ||    | :|    |:
Db     344 lgsqgagggagaaaaaaggagqgggygglgsqgagrgggggagaaaaaaggagqgggyggglgs 403

Qy     179 EGTTNADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQIL-------KAIVHAAKDGG 231
           :|      | |  : ||  |     || |||| |  :|:  |         |   ||  ||
Db     404 qgagqggyggglgsqgagrgglgggqgagaaaaaaaggagqgglgsqgagggagaaaaaagg 463

Qy     232 EKQ------GKKAADRTNPIDAAIGGAGDNDAAAAFATMKK------DDQIAAAMVLRGM 279
             |      | : | |        || |   ||||     :        | |       |:
Db     464 agqggyggglgsqgagr--------gggagaaaaaaggagqggygglgsqgagqggyggl 515

Qy     280 AKD--------GQFALKDAAAAHEGTVKNAVDIIKAAAEAASAASAATGSA------AIG 325
                   || |   ||||  |  :  :   : | : | ||:|| | |        :|
Db     516 gsqgagrgglggqgagaaaaaaaggagqggglg-sqgagggagaaaaaaggagqggyggglg 574

Qy     326 DVVNGNGATAKGGDAKSVNGIAKG 349
               | |    |  | :  |  :|
Db     575 sqgagrgggagaaaaaaggagqg 598


RESULT  14
AAR99055
ID   AAR99055 standard; Protein; 606 AA.
XX
AC   AAR99055;
XX
DT   17-JAN-1997  (first entry)
XX
DE   Spider dragline variant, DP-1B.9 polymer.
XX
KW   Spider; dragline protein; variant;  monomer; polymer;
KW   fibre forming region; Spidroin 1; Nephila clavipes; DP1; mimic;
KW   DP-1A analogue; fibre; high tensile strength; elasticity; clothing;
KW   rope; surgical suture; implant; reinforcement; film; coating.
XX
OS   Synthetic.
XX
PN   WO9429450-A2.
XX
PD   22-DEC-1994.
XX
PF   15-JUN-1994;   94WO-US06689.
XX
PR   15-JUN-1993;   93US-0077600.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Fahnestock SR;
XX
DR   WPI; 1995-036479/05.
XX
PT   New synthetic variants of spider dragline protein - for making
PT   fibres useful as clothing, surgical silk, plastic reinforcement
PT   etc., also related DNA, vectors and transformed cells
XX
PS   Disclosure; Page 88-90; 168pp; English.
XX
CC   This sequence represents a synthetic spider dragline variant polymer,
CC   DP-1B.9.  The sequence of the DP-1B.9 monomer is given in AAR99054.
CC   The polypeptide monomer is a variant based on a consensus sequence
CC   derived from the fibre forming regions of spider dragline protein,
CC   esp. the natural protein 1 (Spidroin 1) from Nephila clavipes.  The
CC   DNA sequence encoding the monomer may be used in the recombinant
CC   production of the variant protein in a recombinant host, e.g. E. coli
CC   or Bacillus subtilis.  Synthetic analogues of DP1 were designed to mimic
CC   the repeating consensus sequence of the natural protein and the pattern
CC   of variation among individual repeats.  This monomer exhibits all of the
CC   regularities of (1)-(5) below.  In addition, it exhibits a regularity of
CC   the natural sequence which is not shared by DP-1A, namely that a repeat
CC   in which both GYG and GRG are deleted is generally preceded by a repeat
CC   lacking the entire poly-alanine repeat, with one intervening repeat.
CC   The sequence of DP-1B matches the natural sequence more closely over
CC   a more extended segment than does DP-1A.  The individual repeats differ
CC   from the consensus sequence given in AAW06201 according to the pattern:
CC          (1) the poly-alanine sequence varies in length from 0-7
CC          residues; (2) when the entire poly-alanine sequence is deleted,
CC   so also is the surrounding sequence encompassing AGRGGLGGQGAGAnGG;
CC          (3) aside from the poly-alanine sequence, deletions usually
CC   encompass integral multiples of three consecutive residues;
CC          (4) deletion of GYG is generally accompanied by deletion of GRG
CC   in the same sequence; and
CC          (5) a repeat in which the entire poly-alanine sequence is
CC   deleted is generally preceded by a repeat containing six alanine
CC   residues.
CC   The proteins may be used to produce fibres of high tensile strength and
CC   elasticity, suitable for clothing, rope, surgical sutures, biomaterials
CC   for implants, plastic reinforcements, films, coatings, etc.
XX
SQ   Sequence   606 AA;


  Query Match             11.9%;  Score 203.5;  DB 16;  Length 606;
  Best Local Similarity   26.5%;  Pred. No. 1.4e-06;
  Matches  101;  Conservative   32;  Mismatches  183;  Indels   65;  Gaps    9;

Qy      25 AASAASAATGNAAIGDVVKNSGAAAKGGEAASVNGIAKGIKGIV---------DAAGKAD 75
           | :||:|| | |  | :         |  ||:  |  :|  | :           || |
Db      25 agaaaaaaaggagqggglgsqgagggagaaaaaaggagqgggyggglgsqgagrgggagaaa 84

Qy      76 AKEGKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGDVVN 135
           |  |     |  |  : ||:         |      |   |  ||:|| | |  | : :
Db      85 aaaggagqggyggglgsqgagqggyggglgsqgagrgglgggagaaaaaaaggagqggglgs 144
```

```
Qy    136 GDVAKAKGGDAASVNGIAKGIKGIVD--------------AAEKADAKEGKLNAAGAEGT 181
Db    145 qgagqggagaaaaaaggagqggygglgsqgagrgggqgagaaaaaaggagqggygglgsqga 204

Qy    182 TNADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQIL-------KAIVHAAKDGGEKQ 234
Db    205 gqggygglgsqgagrgglgggqgagaaaaaaggagqgglgsqgagqgagaaaaaaggagq 264

Qy    235 ------GKKAADRTNPIDAAIGGAGDNDAAAAFATMKK------DDQIAAAMVLRGMAKD 282
Db    265 ggygglgsqgagr--------gggqgagaaaaaaggagqggygglgsqgagqggygglgsq 316

Qy    283 --------GQFALKDAAAAHEGTVKNAVDIIKAAAEAASAASAATGSA------AIGDVV 328
Db    317 gagrgglggqgagaaaaaaaggagqgglg-sqgagggagaaaaaaggagqggygglgsqg 375

Qy    329 NGNGATAKGGDAKSVNGIAKG 349
Db    376 agrgggqgagaaaaaaggagqg 396
```

RESULT 15
AAY40101
ID AAY40101 standard; protein; 606 AA.
XX
AC AAY40101;
XX
DT 19-NOV-1999 (first entry)
XX
DE Polymer of an analogue of spider silk protein spidroine major 1.
XX
KW Spider silk protein; spidroine major 1; cosmetic; make-up;
KW dermatological compositions; hair care; skin care; sunscreen;
KW hormone; moisturizer; skin disorder; skin disorder.
XX
OS Synthetic.
OS Nephila clavipes.
XX
FH Key Location/Qualifiers
FT Peptide 1..101
FT /note= "monomer unit"
XX
PN FR2774588-A1.
XX
PD 13-AUG-1999.
XX
PF 11-FEB-1998; 98FR-0001614.
XX
PR 11-FEB-1998; 98FR-0001614.
XX
PA (OREA) L'OREAL SA.
XX
PI Philippe M, Garson JC, Arraudeau JP;
XX
DR WPI; 1999-510729/43.
XX
PT Cosmetic or dermatological composition containing spider silk protein,
PT for hair or skin care, in make-up or sunscreens -
XX
PS Claim 8; Fig 5B; 32pp; French.
XX
CC The present sequence represents a polymer of an analogue of the spider
CC silk protein spidroine major 1. The protein improves the moisturizing/
CC softening action of the compositions. The protein, and its fragments
CC are used in cosmetic or dermatological compositions. These compositions
CC have use as hair or skin care products; and make-up or sunscreens.
CC As the protein is a good, persistent film-formers on the skin
CC of low surface density, it can be used for delivery of active
CC agents that are generally difficult to administer, e.g. vitamins,
CC hormones, moisturizers or agents for treating disorders of the
CC skin and hair.
XX
SQ Sequence 606 AA;

Query Match 11.9%; Score 203.5; DB 20; Length 606;
Best Local Similarity 26.5%; Pred. No. 1.4e-06;
Matches 101; Conservative 32; Mismatches 183; Indels 65; Gaps 9;

```
Qy     25 AASAASAATGNAAIGDVVKNSGAAAKGGEAASVNGIAKGIKGIV---------DAAGKAD 75
Db     27 agaaaaaaaggagqggglgsqgagqgagaaaaaaggagqgggyggglgsqgagrgggqgagaaa 86

Qy     76 AKEGKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGDVVN 135
Db     87 aaaggagqggygglgsqgagqggyggglgsqgagrgglgggqgagaaaaaaggagqggglgs 146

Qy    136 GDVAKAKGGDAASVNGIAKGIKGIVD--------------AAEKADAKEGKLNAAGAEGT 181
Db    147 qgagqgagaaaaaaggagqggyggglgsqgagrgggqgagaaaaaaggagqggyggglgsqga 206

Qy    182 TNADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQIL-------KAIVHAAKDGGEKQ 234
Db    207 gqggygglgsqgagrgglgggqgagaaaaaaggagqgglgsqgagqgagaaaaaaggagq 266

Qy    235 ------GKKAADRTNPIDAAIGGAGDNDAAAAFATMKK------DDQIAAAMVLRGMAKD 282
Db    267 ggygglgsqgagr--------gggqgagaaaaaaggagqggygglgsqgagqggygglgsq 318

Qy    283 --------GQFALKDAAAAHEGTVKNAVDIIKAAAEAASAASAATGSA------AIGDVV 328
Db    319 gagrgglggqgagaaaaaaggagqgglg-sqgagqgagaaaaaaggagqggygglgsqg 377

Qy    329 NGNGATAKGGDAKSVNGIAKG 349
Db    378 agrgggqgagaaaaaaggagqg 398
```

Search completed: January 10, 2002, 14:05:14
Job time: 168 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 10, 2002, 14:02:26 ; Search time 19.72 Seconds
(without alignments)
398.258 Million cell updates/sec

Title: US-09-445-803-2
Perfect score: 1708
Sequence: 1 KNNDHDNHKGTVKNAVDMAK.....GNGATAKGGDAKSVNGIAKG 349

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	220.5	12.9	738	3	US-08-864-038A-3	Sequence 3, Appli
2	207.5	12.1	606	4	US-09-247-806-6	Sequence 6, Appli
3	203.5	11.9	604	4	US-08-556-978B-63	Sequence 63, Appl
4	203.5	11.9	606	4	US-08-556-978B-23	Sequence 23, Appl
5	203.5	11.9	606	4	US-09-247-806-8	Sequence 8, Appli
6	197	11.5	651	4	US-08-556-978B-19	Sequence 19, Appl
7	197	11.5	651	4	US-09-247-806-1	Sequence 1, Appli
8	197	11.5	718	1	US-08-425-069-2	Sequence 2, Appli
9	197	11.5	718	2	US-08-317-844B-2	Sequence 2, Appli
10	197	11.5	747	3	US-09-034-177-3	Sequence 3, Appli
11	196.5	11.5	832	1	US-08-209-747-2	Sequence 2, Appli
12	196.5	11.5	832	1	US-08-458-298-2	Sequence 2, Appli
13	194.5	11.4	606	4	US-08-556-978B-21	Sequence 21, Appl
14	194.5	11.4	606	4	US-09-247-806-4	Sequence 4, Appli
15	193	11.3	344	1	US-07-941-523-24	Sequence 24, Appl
16	170.5	10.0	1038	1	US-07-609-716-36	Sequence 36, Appl
17	170.5	10.0	1038	4	US-08-475-411A-36	Sequence 36, Appl
18	170.5	10.0	1038	4	US-08-478-029A-36	Sequence 36, Appl
19	163	9.5	235	2	US-08-529-190B-1	Sequence 1, Appli
20	163	9.5	1018	1	US-08-089-862-11	Sequence 11, Appl
21	163	9.5	1018	1	US-08-587-333-18	Sequence 18, Appl
22	163	9.5	1018	5	PCT-US94-07776-16	Sequence 16, Appl
23	163	9.5	1338	2	US-08-728-470-9	Sequence 9, Appli
24	163	9.5	1338	4	US-08-719-641-9	Sequence 9, Appli
25	163	9.5	1529	2	US-08-728-470-10	Sequence 10, Appl
26	163	9.5	1529	4	US-08-719-641-10	Sequence 10, Appl
27	163	9.5	1599	2	US-08-617-697-9	Sequence 9, Appli
28	163	9.5	1600	2	US-08-617-697-10	Sequence 10, Appl
29	162	9.5	766	1	US-08-175-155-53	Sequence 53, Appl
30	162	9.5	766	1	US-08-477-509B-88	Sequence 88, Appl
31	162	9.5	766	2	US-08-707-237A-61	Sequence 61, Appl
32	162	9.5	766	3	US-08-482-085B-88	Sequence 88, Appl
33	162	9.5	979	1	US-08-477-509B-89	Sequence 89, Appl
34	162	9.5	979	3	US-08-482-085B-89	Sequence 89, Appl
35	162	9.5	1050	1	US-08-175-155-54	Sequence 54, Appl
36	161.5	9.5	595	1	US-08-425-069-4	Sequence 4, Appli
37	161.5	9.5	595	2	US-08-317-844B-4	Sequence 4, Appli
38	161	9.4	2123	4	US-08-968-685A-10	Sequence 10, Appl
39	160	9.4	1177	1	US-07-609-716-31	Sequence 31, Appl
40	160	9.4	1177	1	US-08-175-155-29	Sequence 29, Appl
41	160	9.4	1177	1	US-08-477-509B-64	Sequence 64, Appl
42	160	9.4	1177	2	US-08-707-237A-35	Sequence 35, Appl
43	160	9.4	1177	3	US-08-482-085B-64	Sequence 64, Appl
44	160	9.4	1177	4	US-08-475-411A-31	Sequence 31, Appl
45	160	9.4	1177	4	US-08-478-029A-31	Sequence 31, Appl

ALIGNMENTS

```
RESULT 1
US-08-864-038A-3
; Sequence 3, Application US/08864038A
; Patent No. 6001592
;  GENERAL INFORMATION:
;    APPLICANT:  Kunio NAKASHIMA et al.
;    TITLE OF INVENTION:  NOVEL POLYPEPTIDE GENE cDNA, VECTOR
;    TITLE OF INVENTION:  CONTAINING SAID cDNA, HOST CELLS TRANSFORMED WITH SAID
;    TITLE OF INVENTION:  VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
;    TITLE OF INVENTION:  SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBOD
;    TITLE OF INVENTION:  TO SAID POLYPEPTIDE
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  812-5 Hirano
;      STREET:  Isshinden
;      CITY:  Tsu-city
;      STATE:  Mie-prefecture
;      COUNTRY:  JAPAN
;      ZIP:  514-01
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 MB storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Microsoft Windows 95
;      SOFTWARE:  Word Perfect 6.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/864,038A
;      FILING DATE:  May 28, 1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 8-184459
;      FILING DATE:  15-July-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  C. Bruce Hamburg
;      REGISTRATION NUMBER:  22,389
;      REFERENCE/DOCKET NUMBER:  F-5610
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)986-2340
;      TELEFAX:  (212)953-7733
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  738
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Pinctada fucata
;      CELL TYPE:  mantle epithelial cell
;    FEATURE:
;      NAME/KEY:  peptide
;      LOCATION:  from 1 to 738
;      IDENTIFICATION METHOD:  E (by experiment)
```

US-08-864-038A-3

Query Match 12.9%; Score 220.5; DB 3; Length 738;
Best Local Similarity 29.6%; Pred. No. 2.2e-11;
Matches 113; Conservative 29; Mismatches 167; Indels 73; Gaps 14;

```
Qy      19 AKAAEEAASAASAATGNAAIGDVVKNSGAAAKGGEAASVNGIAKGIKGIVDAAGKADAKE 78
Db     324 AAAAAAAAAAAGGGGGGGGGGGGGGAGAAAAAAAAAASASASRQMSGIRDALG--DIKD 381

Qy      79 GKLDATGAEGTTNVNAGKL-FVKRAADDGGDA--DDAG------KAAAAVAASAATGNAA 129
Db     382 -LLRSNGASAKASAKASAVASTKSQIDDLKDVLKDLAGLLKSSASASASASASASAGGGG 440

Qy     130 IGDVVNGDVAKAKGGDAASVNGI-AKGIKGIVDAAEKADAKEGKLNAAGAEGTTNADAGK 188
Db     441 GGGNGGGNGGGGGGGAGALAAALAAAGAGGGLGGGGGGGALAAALAAAGAGGGGFGGLGG 500

Qy     189 LFVKNAGNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAADRTNPIDAA 248
Db     501 L-----GGLGGGSAAAAAAAAAAASGGGGRALRRALRRQMRGG---GSAAAAAAAAAAAA 552

Qy     249 IGG---------------------AGDNDAAAAFATMKKDDQIAAAMVL---------RG 278
Db     553 GGGWGGGMGGGFGVGLGGGFGGGFGGGSSAAAAAAA-------AAAAGFGGGGRRGRGRG 605

Qy     279 MAKDGQ----FALKDAAAAHEGTVKNAVDIIKAAAEAASAASAATGSAAIGDVV------ 328
Db     606 RGGDGDGNGASAVAAAAAAAAAAGGSAADV----AAAAAAAAAMYGDGADGPDFDNGFGG 661

Qy     329 -NGNGATAKGGDAKSVNGIAKG 349
Db     662 GNGNGGGGSGGGGSGGGGSGGG 683
```

RESULT 2
US-09-247-806-6
; Sequence 6, Application US/09247806
; Patent No. 6280747
; GENERAL INFORMATION:
; APPLICANT: PHILLIPPE, Michel
; APPLICANT: GARSON, Jean-Claude
; APPLICANT: ARRAUDEAU, Jean-Pierre
; TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT
; TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN
; TITLE OF INVENTION: ANALOG
; FILE REFERENCE: 6388-0365-0
; CURRENT APPLICATION NUMBER: US/09/247,806
; CURRENT FILING DATE: 1999-02-11
; EARLIER APPLICATION NUMBER: FR 98/01614
; EARLIER FILING DATE: 1998-02-11
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 606
; TYPE: PRT
; ORGANISM: Nephila clavipes
US-09-247-806-6

Query Match 12.1%; Score 207.5; DB 4; Length 606;
Best Local Similarity 26.8%; Pred. No. 2.3e-10;
Matches 102; Conservative 32; Mismatches 182; Indels 65; Gaps 9;

```
Qy      25 AASAASAATGNAAIGDVVKNSGAAAKGGEAASVNGIAKGIKGIV---------DAAGKAD 75
Db     229 AGAAAAAAAGGAGQGGLGSQGAGQGAGAAAAAAGGAGQGGYGGLGSQGAGRGGQGAGAAA 288

Qy      76 AKEGKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGDVVN 135
Db     289 AAAGGAGQGGYGGLGSQGAGQGGYGGLGSQGAGRGGLGGQGAAAAAAAAAGGAGQGGLGS 348

Qy     136 GDVAKAKGGDAASVNGIAKGIKGIVD----------AAEKADAKEGKLNAAGAEGT 181
Db     349 QGAGQGAGAAAAAAGGAGQGGYGGLGSQGAGRGGQGAGAAAAAAGGAGQGGYGGLGSQGA 408

Qy     182 TNADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQIL-------KAIVHAAKDGGEKQ 234
Db     409 GQGGYGGLGSQGAGRGGLGGQGAGAAAAAAAGGAGQGGLGSQGAGQGAGAAAAAAGGAGQ 468

Qy     235 ------GKKAADRTNPIDAAIGGAGDNDAAAAFATMKK------DDQIAAAMVLRGMAKD 282
Db     469 GGYGGLGSQGAGR--------GGQGAGAAAAAAGGAGQGGYGGLGSQGAGQGGYGGLGSQ 520

Qy     283 --------GQFALKDAAAAHEGTVKNAVDIIKAAAEAASAASAATGSA------AIGDVV 328
Db     521 GAGRGGLGGQGAGAAAAAAAGGAGQGGLG-SQGAGQGAGAAAAAAGGAGQGGYGGLGSQG 579

Qy     329 NGNGATAKGGDAKSVNGIAKG 349
Db     580 AGRGGQGAGAAAAAAGGAGQG 600
```

RESULT 3
US-08-556-978B-63
; Sequence 63, Application US/08556978B
; Patent No. 6268169
; GENERAL INFORMATION:
; APPLICANT: FAHNESTOCK, STEPHEN F.
; TITLE OF INVENTION: NOVEL RECOMBINANTLY PRODUCED
; TITLE OF INVENTION: SPIDER SILK ANALOGS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 95
; SOFTWARE: MICROSOFT WORD FOR WINDOWS 95
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/556,978B
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/077,600
; FILING DATE: JUNE 15, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: FLOYD, LINDA AXAMETHY
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: CR-9389-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-8112
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 604 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-556-978B-63

Query Match 11.9%; Score 203.5; DB 4; Length 604;
Best Local Similarity 26.6%; Pred. No. 5.2e-10;
Matches 102; Conservative 35; Mismatches 162; Indels 85; Gaps 13;

```
Qy      19 AKAAEEAASAASAATGNA------AIGDVVKNSGAAAKGGEAASVNGIAKGIKGIVDAAG 72
Db     247 SQGAGQGAGAAAAAAGGAGQGGYGGLGSQGAGRGGQGAGAAAAAAGGAGGGYGGLGSGAG 306
```

```
Qy      73 KADAKEGKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGD 132
           :|     |::|     ||:          ||      |   |  ||:|| | |  |
Db     307 -----QGGYGGLGSQG-----AGR---------GG----LGGQGAGAAAAAAAGGAGQGG 343

Qy     133 VVNGDVAKAKGGDAASVNGIAKGIKGIVD--------------AAEKADAKEGKLNAAGA 178
           : :    :  |  ||:  |  :|  | :                 ||    | :|     |:
Db     344 LGSQGAGQGAGAAAAAAGGAGQGGYGGLGSQGAGRGGQGAGAAAAAAGGAGQGGYGGLGS 403

Qy     179 EGTTNADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQIL-------KAIVHAAKDGG 231
           :|      | |  : ||  |     || |||| |  :|:  |         |   ||  ||
Db     404 QGAGQGGYGGLGSQGAGRGGLGGQGAGAAAAAAAGGAGQGGLGSQGAGQGAGAAAAAAGG 463

Qy     232 EKQ------GKKAADRTNPIDAAIGGAGDNDAAAAFATMKK------DDQIAAAMVLRGM 279
           |       | : | |         || |  ||||     :       | |       |:
Db     464 AGQGGYGGLGSQGAGR--------GGQGAGAAAAAAGGAGQGGYGGLGSQGAGQGGYGGL 515

Qy     280 AKD--------GQFALKDAAAAHEGTVKNAVDIIKAAAEAASAASAATGSA------AIG 325
                  || |   ||||  |  :  :   : | : | ||:|| | |       :|
Db     516 GSQGAGRGGLGGQGAGAAAAAAAGGAGQGGLG-SQGAGQGAGAAAAAAGGAGQGGYGGLG 574

Qy     326 DVVNGNGATAKGGDAKSVNGIAKG 349
           | |    |  | :  |  :|
Db     575 SQGAGRGGQGAGAAAAAAGGAGQG 598
```

RESULT 4
US-08-556-978B-23
; Sequence 23, Application US/08556978B
; Patent No. 6268169
; GENERAL INFORMATION:
; APPLICANT: FAHNESTOCK, STEPHEN F.
; TITLE OF INVENTION: NOVEL RECOMBINANTLY PRODUCED
; TITLE OF INVENTION: SPIDER SILK ANALOGS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 95
; SOFTWARE: MICROSOFT WORD FOR WINDOWS 95
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/556,978B
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/077,600
; FILING DATE: JUNE 15, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: FLOYD, LINDA AXAMETHY
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: CR-9389-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-8112
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 606 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-556-978B-23

Query Match 11.9%; Score 203.5; DB 4; Length 606;
Best Local Similarity 26.5%; Pred. No. 5.2e-10;
Matches 101; Conservative 32; Mismatches 183; Indels 65; Gaps 9;

```
Qy      25 AASAASAATGNAAIGDVVKNSGAAAKGGEAASVNGIAKGIKGIV---------DAAGKAD 75
           | :||:|| | |  | :          |  ||:  |  :|  | :           || |
Db      25 AGAAAAAAAGGAGQGGLGSQGAGQGAGAAAAAAGGAGQGGYGGLGSQGAGRGGQGAGAAA 84

Qy      76 AKEGKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGDVVN 135
           |  |      |  |  :  ||:          |       |   |  ||:|| | |  | : :
Db      85 AAAGGAGQGGYGGLGSQGAGQGGYGGLGSQGAGRGGLGGQGAGAAAAAAAGGAGQGGLGS 144

Qy     136 GDVAKAKGGDAASVNGIAKGIKGIVD--------------AAEKADAKEGKLNAAGAEGT 181
               :  |  ||:  |  :|  | :                ||     | :|     |::|
Db     145 QGAGQGAGAAAAAAGGAGQGGYGGLGSQGAGRGGQGAGAAAAAAGGAGQGGYGGLGSQGA 204

Qy     182 TNADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQIL-------KAIVHAAKDGGEKQ 234
               | |  : ||  |      || |||| |  :|:  |         |   ||  ||  |
Db     205 GQGGYGGLGSQGAGRGGLGGQGAGAAAAAAAGGAGQGGLGSQGAGQGAGAAAAAAGGAGQ 264

Qy     235 ------GKKAADRTNPIDAAIGGAGDNDAAAAFATMKK------DDQIAAAMVLRGMAKD 282
                  | : | |         || |   ||||     :       | |       |:
Db     265 GGYGGLGSQGAGR--------GGQGAGAAAAAAGGAGQGGYGGLGSQGAGQGGYGGLGSQ 316

Qy     283 --------GQFALKDAAAAHEGTVKNAVDIIKAAAEAASAASAATGSA------AIGDVV 328
                     || |   ||||  |  :  :   : | : | ||:|| | |       :|
Db     317 GAGRGGLGGQGAGAAAAAAAGGAGQGGLG-SQGAGQGAGAAAAAAGGAGQGGYGGLGSQG 375

Qy     329 NGNGATAKGGDAKSVNGIAKG 349
           | |    |  | :  |  :|
Db     376 AGRGGQGAGAAAAAAGGAGQG 396
```

RESULT 5
US-09-247-806-8
; Sequence 8, Application US/09247806
; Patent No. 6280747
; GENERAL INFORMATION:
; APPLICANT: PHILLIPPE, Michel
; APPLICANT: GARSON, Jean-Claude
; APPLICANT: ARRAUDEAU, Jean-Pierre
; TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT
; TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN
; TITLE OF INVENTION: ANALOG
; FILE REFERENCE: 6388-0365-0
; CURRENT APPLICATION NUMBER: US/09/247,806
; CURRENT FILING DATE: 1999-02-11
; EARLIER APPLICATION NUMBER: FR 98/01614
; EARLIER FILING DATE: 1998-02-11
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 606
; TYPE: PRT
; ORGANISM: Nephila clavipes
US-09-247-806-8

Query Match 11.9%; Score 203.5; DB 4; Length 606;
Best Local Similarity 26.5%; Pred. No. 5.2e-10;
Matches 101; Conservative 32; Mismatches 183; Indels 65; Gaps 9;

```
Qy      25 AASAASAATGNAAIGDVVKNSGAAAKGGEAASVNGIAKGIKGIV---------DAAGKAD 75
           | :||:|| | |  | :          |  ||:  |  :|  | :           || |
Db      25 AGAAAAAAAGGAGQGGLGSQGAGQGAGAAAAAAGGAGQGGYGGLGSQGAGRGGQGAGAAA 84

Qy      76 AKEGKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGDVVN 135
           |  |      |  |  :  ||:          |       |   |  ||:|| | |  | : :
Db      85 AAAGGAGQGGYGGLGSQGAGQGGYGGLGSQGAGRGGLGGQGAGAAAAAAAGGAGQGGLGS 144

Qy     136 GDVAKAKGGDAASVNGIAKGIKGIVD--------------AAEKADAKEGKLNAAGAEGT 181
               :  |  ||:  |  :|  | :                ||     | :|     |::|
Db     145 QGAGQGAGAAAAAAGGAGQGGYGGLGSQGAGRGGQGAGAAAAAAGGAGQGGYGGLGSQGA 204

Qy     182 TNADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQIL-------KAIVHAAKDGGEKQ 234
```

```
Db    205 GQGGYGGLGSQGAGRGGLGGQGAGAAAAAAAGGAGQGGLGSQGAGQGAGAAAAAAGGAGQ 264

Qy    235 ------GKKAADRTNPIDAAIGGAGDNDAAAAFATMKK------DDQIAAAMVLRGMAKD 282
Db    265 GGYGGLGSQGAGR--------GGQGAGAAAAAAGGAGQGGYGGLGSQGAGQGGYGGLGSQ 316

Qy    283 --------GQFALKDAAAAHEGTVKNAVDIIKAAAEAASAASAATGSA------AIGDVV 328
Db    317 GAGRGGLGGQGAGAAAAAAAGGAGQGGLG-SQGAGQGAGAAAAAAGGAGQGGYGGLGSQG 375

Qy    329 NGNGATAKGGDAKSVNGIAKG 349
Db    376 AGRGGQGAGAAAAAAGGAGQG 396


RESULT   6
US-08-556-978B-19
; Sequence 19, Application US/08556978B
; Patent No. 6268169
;  GENERAL INFORMATION:
;    APPLICANT:  FAHNESTOCK, STEPHEN F.
;    TITLE OF INVENTION:  NOVEL RECOMBINANTLY PRODUCED
;    TITLE OF INVENTION:  SPIDER SILK ANALOGS
;    NUMBER OF SEQUENCES:  107
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E. I. DU PONT DE NEMOURS AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINTON
;      STATE:  DELAWARE
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.50 INCH
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MICROSOFT WINDOWS 95
;      SOFTWARE:  MICROSOFT WORD FOR WINDOWS 95
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/556,978B
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/077,600
;      FILING DATE:  JUNE 15, 1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FLOYD, LINDA AXAMETHY
;      REGISTRATION NUMBER:  33,692
;      REFERENCE/DOCKET NUMBER:  CR-9389-A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  302-892-8112
;      TELEFAX:  302-773-0164
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  651 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-556-978B-19


  Query Match             11.5%;  Score 197;  DB 4;  Length 651;
  Best Local Similarity   26.1%;  Pred. No. 2.1e-09;
  Matches   98;  Conservative   32;  Mismatches  164;  Indels   82;  Gaps   13;

Qy      20 KAAEEAASAASAATGNAAIGDV-VKNSGAAAKGGE-----AASVNGIAKGIKGIVDAAGK 73
Db    234 QGAGQGAGASAAAAGGAGQGGYGGLGSQGAGRGGEGAGAAAAAAGGAGQGGYGGLGGQGA 293

Qy      74 ADAKEGKLDATGA--EGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASA-------- 123
Db    294 GQGGYGGLGSQGAGRGGLGGQGAGAAAAGGAGQGGLGGQGAGQGAGAAAAAAGGAGQGGY 353

Qy    124 ------ATGNAAIGDVVNGDVAKAKGGDAA-------SVNGIAKGIKGIVDAAEKADAKE 170
Db    354 GGLGSQGAGRGGLGGQGAGAVAAAAAGGAGQGGYGGLGSQGAGRGGQG----AGAAAAAA 409

Qy    171 GKLNAAGAEGTTNADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDG 230
Db    410 GGAGQRGYGGLGNQGAGR------GGLGGQG--AGAAAAAAAGGAGQGGYGGL---GNQG 458

Qy    231 GEKQGKKAADRTNPIDAAIGGAGDN----------------DAAAAFATMKKDDQIAAA 273
Db    459 AGRGGQGAA-------AAAGGAGQGGYGGLGSQGAGRGGQGAGAAAAAAVGAGQEGIRG- 510

Qy    274 MVLRGMAKDGQFALKDAAAAHEGTVKNAVDIIKAAAEAASAASAATGSAAIGDVVNGNGA 333
Db    511 ---QGAGQGGYGGLGSQGSGRGG----------LGGQGAGAAAAAAGGAGQGGLGGQGAG 557

Qy    334 TAKGGDAKSVNGIAKG 349
Db    558 QGAGAAAAAAGGVRQG 573


RESULT   7
US-09-247-806-1
; Sequence 1, Application US/09247806
; Patent No. 6280747
; GENERAL INFORMATION:
;  APPLICANT: PHILLIPPE, Michel
;  APPLICANT: GARSON, Jean-Claude
;  APPLICANT: ARRAUDEAU, Jean-Pierre
;  TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT
;  TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN
;  TITLE OF INVENTION: ANALOG
;  FILE REFERENCE: 6388-0365-0
;  CURRENT APPLICATION NUMBER: US/09/247,806
;  CURRENT FILING DATE: 1999-02-11
;  EARLIER APPLICATION NUMBER: FR 98/01614
;  EARLIER FILING DATE: 1998-02-11
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 651
;   TYPE: PRT
;   ORGANISM: Nephila clavipes
US-09-247-806-1


  Query Match             11.5%;  Score 197;  DB 4;  Length 651;
  Best Local Similarity   26.1%;  Pred. No. 2.1e-09;
  Matches   98;  Conservative   32;  Mismatches  164;  Indels   82;  Gaps   13;

Qy      20 KAAEEAASAASAATGNAAIGDV-VKNSGAAAKGGE-----AASVNGIAKGIKGIVDAAGK 73
Db    234 QGAGQGAGASAAAAGGAGQGGYGGLGSQGAGRGGPGAGAAAAAAGGAGQGGYGGLGGQGA 293

Qy      74 ADAKEGKLDATGA--EGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASA-------- 123
Db    294 GQGGYGGLGSQGAGRGGLGGQGAGAAAAGGAGQGGLGGQGAGQGAGAAAAAAGGAGQGGY 353

Qy    124 ------ATGNAAIGDVVNGDVAKAKGGDAA-------SVNGIAKGIKGIVDAAEKADAKE 170
Db    354 GGLGSQGAGRGGLGGQGAGAVAAAAAGGAGQGGYGGLGSQGAGRGGQG----AGAAAAAA 409

Qy    171 GKLNAAGAEGTTNADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDG 230
Db    410 GGAGQRGYGGLGNQGAGR------GGLGGQG--AGAAAAAAAGGAGQGGYGGL---GNQG 458

Qy    231 GEKQGKKAADRTNPIDAAIGGAGDN----------------DAAAAFATMKKDDQIAAA 273
Db    459 AGRGGQGAA-------AAAGGAGQGGYGGLGSQGAGRGGQGAGAAAAAAVGAGQEGIRG- 510

Qy    274 MVLRGMAKDGQFALKDAAAAHEGTVKNAVDIIKAAAEAASAASAATGSAAIGDVVNGNGA 333
Db    511 ---QGAGQGGYGGLGSQGSGRGG----------LGGQGAGAAAAAAGGAGQGGLGGQGAG 557
```

```
Qy    334 TAKGGDAKSVNGIAKG 349
          |  | :  |: :|
Db    558 QGAGAAAAAAGGVRQG 573


RESULT   8
US-08-425-069-2
; Sequence 2, Application US/08425069
; Patent No. 5728810
;  GENERAL INFORMATION:
;    APPLICANT:  Lewis, Randolph V.
;    APPLICANT:  Xu, Ming
;    APPLICANT:  Hinman, Michael B.
;    TITLE OF INVENTION:  ISOLATED DNA CODING FOR SPIDER SILK
;    TITLE OF INVENTION:  PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
;    TITLE OF INVENTION:  CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
;    NUMBER OF SEQUENCES:  69
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;      STREET:  301 No. 5728810th Washington Street
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22046
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/425,069
;      FILING DATE:  19-APR-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy Jr., Gerald M
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1447-106P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 205-8000
;      TELEFAX:  (703) 205-8050
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  718 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-425-069-2

  Query Match             11.5%;  Score 197;  DB 1;  Length 718;
  Best Local Similarity   26.1%;  Pred. No. 2.4e-09;
  Matches   98;  Conservative   32;  Mismatches  164;  Indels   82;  Gaps   13;

Qy     20 KAAEEAASAASAATGNAAIGDV-VKNSGAAAKGGE-----AASVNGIAKGIKGIVDAAGK 73
          : | : | |::|| | |  |       |  | :|||     ||:  |  :|  | :   |
Db    234 QGAGQGAGASAAAAGGAGQGGYGGLGSQGAGRGGEGAGAAAAAAGGAGQGGYGGLGGQGA 293

Qy     74 ADAKEGKLDATGA--EGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASA-------- 123
             | | : ||   |     ||       |   |      ||: | | ||:|
Db    294 GQGGYGGLGSQGAGRGGLGGQGAGAAAAGGAGQGGLGGQGAGQGAGAAAAAAGGAGQGGY 353

Qy    124 ------ATGNAAIGDVVNGDVAKAKGGDAA-------SVNGIAKGIKGIVDAAEKADAKE 170
                |   :|    | || |  | |           |  :| :|    |  | |
Db    354 GGLGSQGAGRGGLGGQGAGAVAAAAAGGAGQGGYGGLGSQGAGRGGQG----AGAAAAAA 409

Qy    171 GKLNAAGAEGTTNADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDG 230
          |     |  |  ||:      | :||:   || |||| |  :|:     :      |
Db    410 GGAGQRGYGGLGNQGAGR------GGLGGQG--AGAAAAAAAGGAGQGGYGGL---GNQG 458

Qy    231 GEKQGKKAADRTNPIDAAIGGAGDN----------------DAAAAFATMKKDDQIAAA 273
            : |: ||       || ||||                   ||||  |     : |
Db    459 AGRGGQGAA-------AAAGGAGQGGYGGLGSQGAGRGGQGAGAAAAAAVGAGQEGIRG- 510

Qy    274 MVLRGMAKDGQFALKDAAAAHEGTVKNAVDIIKAAAEAASAASAATGSAAIGDVVNGNGA 333
            :|  : |   |    :   |            : | ||:|| | |  | :
Db    511 ---QGAGQGGYGGLGSQGSGRGG----------LGGQGAGAAAAAAGGAGQGGLGGQGAG 557

Qy    334 TAKGGDAKSVNGIAKG 349
          |  | :  |: :|
Db    558 QGAGAAAAAAGGVRQG 573


RESULT   9
US-08-317-844B-2
; Sequence 2, Application US/08317844B
; Patent No. 5989894
;  GENERAL INFORMATION:
;    APPLICANT:  Lewis, Randolph V.
;    APPLICANT:  Xu, Ming
;    APPLICANT:  Hinman, Michael B.
;    TITLE OF INVENTION:  ISOLATED DNA CODING FOR SPIDER SILK
;    TITLE OF INVENTION:  PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
;    TITLE OF INVENTION:  CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
;    NUMBER OF SEQUENCES:  62
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;      STREET:  301 No. 5989894th Washington Street
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22046
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/317,844B
;      FILING DATE:  04-OCT-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy Jr., Gerald M
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1447-105P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 241-1300
;      TELEFAX:  (703) 241-2848
;      TELEX:  248345
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  718 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-317-844B-2

  Query Match             11.5%;  Score 197;  DB 2;  Length 718;
  Best Local Similarity   26.1%;  Pred. No. 2.4e-09;
  Matches   98;  Conservative   32;  Mismatches  164;  Indels   82;  Gaps   13;

Qy     20 KAAEEAASAASAATGNAAIGDV-VKNSGAAAKGGE-----AASVNGIAKGIKGIVDAAGK 73
          : | : | |::|| | |  |       |  | :|||     ||:  |  :|  | :   |
Db    234 QGAGQGAGASAAAAGGAGQGGYGGLGSQGAGRGGEGAGAAAAAAGGAGQGGYGGLGGQGA 293

Qy     74 ADAKEGKLDATGA--EGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASA-------- 123
             | | : ||   |     ||       |   |      ||: | | ||:|
Db    294 GQGGYGGLGSQGAGRGGLGGQGAGAAAAGGAGQGGLGGQGAGQGAGAAAAAAGGAGQGGY 353

Qy    124 ------ATGNAAIGDVVNGDVAKAKGGDAA-------SVNGIAKGIKGIVDAAEKADAKE 170
                |   :|    | || |  | |           |  :| :|    |  | |
Db    354 GGLGSQGAGRGGLGGQGAGAVAAAAAGGAGQGGYGGLGSQGAGRGGQG----AGAAAAAA 409

Qy    171 GKLNAAGAEGTTNADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDG 230
```

```
Db    410 GGAGQRGYGGLGNQGAGR------GGLGGQG--AGAAAAAAAGGAGQGGYGGL---GNQG 458

Qy    231 GEKQGKKAADRTNPIDAAIGGAGDN----------------DAAAAFATMKKDDQIAAA 273
Db    459 AGRGGQGAA-------AAAGGAGQGGYGGLGSQGAGRGGQGAGAAAAAAVGAGQEGIRG- 510

Qy    274 MVLRGMAKDGQFALKDAAAAHEGTVKNAVDIIKAAAEAASAASAATGSAAIGDVVNGNGA 333
Db    511 ---QGAGQGGYGGLGSQGSGRGG----------LGGQGAGAAAAAAGGAGQGGLGGQGAG 557

Qy    334 TAKGGDAKSVNGIAKG 349
Db    558 QGAGAAAAAAGGVRQG 573


RESULT 10
US-09-034-177-3
; Sequence 3, Application US/09034177
; Patent No. 6127146
;  GENERAL INFORMATION:
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Guegler, Karl J.
;    APPLICANT:  Corley, Neil C.
;    TITLE OF INVENTION:  HUMAN FIBROUS PROTEIN
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/034,177
;      FILING DATE:  HEREWITH
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0486 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-855-0555
;      TELEFAX:  650-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  747 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  GenBank
;      CLONE:  GI 1174414
US-09-034-177-3


  Query Match           11.5%;  Score 197;  DB 3;  Length 747;
  Best Local Similarity  26.1%;  Pred. No. 2.5e-09;
  Matches   98;  Conservative   32;  Mismatches  164;  Indels   82;  Gaps   13;

Qy     20 KAAEEAASAASAATGNAAIGDV-VKNSGAAAKGGE-----AASVNGIAKGIKGIVDAAGK 73
Db    234 QGAGQGAGASAAAAGGAGQGGYGGLGSQGAGRGGEGAGAAAAAAGGAGQGGYGGLGGQGA 293

Qy     74 ADAKEGKLDATGA--EGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASA-------- 123
Db    294 GQGGYGGLGSQGAGRGGLGGQGAGAAAAGGAGQGGLGGQGAGQGAGAAAAAAGGAGQGGY 353

Qy    124 ------ATGNAAIGDVVNGDVAKAKGGDAA-------SVNGIAKGIKGIVDAAEKADAKE 170
Db    354 GGLGSQGAGRGGLGGQGAGAVAAAAAGGAGQGGYGGLGSQGAGRGGQG----AGAAAAAA 409

Qy    171 GKLNAAGAEGTTNADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDG 230
Db    410 GGAGQRGYGGLGNQGAGR------GGLGGQG--AGAAAAAAAGGAGQGGYGGL---GNQG 458

Qy    231 GEKQGKKAADRTNPIDAAIGGAGDN----------------DAAAAFATMKKDDQIAAA 273
Db    459 AGRGGQGAA-------AAAGGAGQGGYGGLGSQGAGRGGQGAGAAAAAAVGAGQEGIRG- 510

Qy    274 MVLRGMAKDGQFALKDAAAAHEGTVKNAVDIIKAAAEAASAASAATGSAAIGDVVNGNGA 333
Db    511 ---QGAGQGGYGGLGSQGSGRGG----------LGGQGAGAAAAAAGGAGQGGLGGQGAG 557

Qy    334 TAKGGDAKSVNGIAKG 349
Db    558 QGAGAAAAAAGGVRQG 573


RESULT 11
US-08-209-747-2
; Sequence 2, Application US/08209747
; Patent No. 5733771
;  GENERAL INFORMATION:
;    APPLICANT:  Lewis, Randolph V.
;    APPLICANT:  Colgin, Mark
;    TITLE OF INVENTION:  cDNAs Encoding Minor Ampullate Spider
;    TITLE OF INVENTION:  Silk Proteins
;    NUMBER OF SEQUENCES:  56
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;      STREET:  P.O. Box 747
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22040-3487
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/209,747
;      FILING DATE:  14-MAR-1994
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy Jr., Gerald M.
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1447-104P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-205-8000
;      TELEFAX:  703-205-8050
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  832 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
;      ORGANISM:  N. clavipes
;      TISSUE TYPE:  minor ampullate gland
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..309
```

```
US-08-209-747-2

  Query Match           11.5%;  Score 196.5;  DB 1;  Length 832;
  Best Local Similarity  28.4%;  Pred. No. 3.2e-09;
  Matches   97;  Conservative   14;  Mismatches  185;  Indels   45;  Gaps   10;

Qy      19 AKAAEEAASAASAATGNAAIGDVVKNSG---AAAKGGEAASVNGIAKGIKGIVDAAGKAD 75
           |  |   |:| : | |    |     :|   |||| |  |    |  :|      ||  |
Db     380 AAGAGAGAAAGAGAGGYGGQGGYGAGAGAGAAAAAGAGAGGAGGYGRGAGAGAGAAAGAG 439

Qy      76 AKEGKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGK-----AAAAVAASAATGNAAI 130
           |        | :|     ||      ||   | |   |:     | ||  | | || |
Db     440 A-----GGYGGQGGYGAGAGAGAAAAAATGAGGAGGYGRGAGAGAGAAAGAGAGTGGAGY 494

Qy     131 GDVVNGDVAKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGAEGTTNADAGKLF 190
           |    |      |  ||: |   |  |        | |  |    |||     | ||
Db     495 GG--QGGYGAGAGAGAAAAAGAGAGGAGY---GRGAGAGAGAAAGAGAGAAAGAGAG--- 546

Qy     191 VKNAGNVGGEAG-DAG-KAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAADRTNPIDAA 248
              ||  ||: |  || :| || || :|          : |    |  ||
Db     547 ---AGGYGGQGGYGAGARAGAAAAAGAGAGGAAGYSRGGRAGAAGAGAGAAAGAGAGAGG 603

Qy     249 IGGAGDNDAAAAFATMKKDDQIAAAMVLRGMAKDGQFALKDAAAAHEGTVKNAVDIIKAA 308
           || |   | |           |||    |    | :     | |  |           |
Db     604 YGGQGGYGAGAGAG--------AAAAAGAGSGGAGGYGRGAGAGAAAG-----------A 644

Qy     309 AEAASAASAATGSAAIGDVVNGNGATAKGGDAKSVNGIAKG 349
           || | : | |    |    | || | |      | :|
Db     645 GAAAGAGAGAGGYGGQGGYGAGAGAAAAAGAGAGRGGYGRG 685


RESULT  12
US-08-458-298-2
; Sequence 2, Application US/08458298
; Patent No. 5756677
;  GENERAL INFORMATION:
;    APPLICANT:  Lewis, Randolph V.
;    APPLICANT:  Colgin, Mark
;    TITLE OF INVENTION:  cDNAs Encoding Minor Ampullate Spider
;    TITLE OF INVENTION:  Silk Proteins
;    NUMBER OF SEQUENCES:  56
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;      STREET:  P.O. Box 747
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22040-3487
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,298
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/209,747
;      FILING DATE:  14-MAR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy Jr., Gerald M.
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1447-104P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-205-8000
;      TELEFAX:  703-205-8050
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  832 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
;      ORGANISM:  N. clavipes
;      TISSUE TYPE:  minor ampullate gland
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..309
US-08-458-298-2


  Query Match           11.5%;  Score 196.5;  DB 1;  Length 832;
  Best Local Similarity  28.4%;  Pred. No. 3.2e-09;
  Matches   97;  Conservative   14;  Mismatches  185;  Indels   45;  Gaps   10;

Qy      19 AKAAEEAASAASAATGNAAIGDVVKNSG---AAAKGGEAASVNGIAKGIKGIVDAAGKAD 75
           |  |   |:| : | |    |     :|   ||| |  |    |  :|      ||  |
Db     380 AAGAGAGAAAGAGAGGYGGQGGYGAGAGAGAAAAAGAGAGGAGGYGRGAGAGAGAAAGAG 439

Qy      76 AKEGKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGK-----AAAAVAASAATGNAAI 130
           |        | :|     ||      ||   | |   |:     | ||  | | || |
Db     440 A-----GGYGGQGGYGAGAGAGAAAAAATGAGGAGGYGRGAGAGAGAAAGAGAGTGGAGY 494

Qy     131 GDVVNGDVAKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGAEGTTNADAGKLF 190
           |    |      |  ||: |   |  |        | |  |    |||     | ||
Db     495 GG--QGGYGAGAGAGAAAAAGAGAGGAGY---GRGAGAGAGAAAGAGAGAAAGAGAG--- 546

Qy     191 VKNAGNVGGEAG-DAG-KAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAADRTNPIDAA 248
              ||  ||: |  || :| || || :|          : |    |  ||
Db     547 ---AGGYGGQGGYGAGARAGAAAAAGAGAGGAAGYSRGGRAGAAGAGAGAAAGAGAGAGG 603

Qy     249 IGGAGDNDAAAAFATMKKDDQIAAAMVLRGMAKDGQFALKDAAAAHEGTVKNAVDIIKAA 308
           || |   | |           |||    |    | :     | |  |           |
Db     604 YGGQGGYGAGAGAG--------AAAAAGAGSGGAGGYGRGAGAGAAAG-----------A 644

Qy     309 AEAASAASAATGSAAIGDVVNGNGATAKGGDAKSVNGIAKG 349
           || | : | |    |    | || | |      | :|
Db     645 GAAAGAGAGAGGYGGQGGYGAGAGAAAAAGAGAGRGGYGRG 685


RESULT  13
US-08-556-978B-21
; Sequence 21, Application US/08556978B
; Patent No. 6268169
;  GENERAL INFORMATION:
;    APPLICANT:  FAHNESTOCK, STEPHEN F.
;    TITLE OF INVENTION:  NOVEL RECOMBINANTLY PRODUCED
;    TITLE OF INVENTION:  SPIDER SILK ANALOGS
;    NUMBER OF SEQUENCES:  107
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E. I. DU PONT DE NEMOURS AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINTON
;      STATE:  DELAWARE
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.50 INCH
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MICROSOFT WINDOWS 95
;      SOFTWARE:  MICROSOFT WORD FOR WINDOWS 95
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/556,978B
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/077,600
;      FILING DATE:  JUNE 15, 1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FLOYD, LINDA AXAMETHY
;      REGISTRATION NUMBER:  33,692
```

```
;      REFERENCE/DOCKET NUMBER:  CR-9389-A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  302-892-8112
;      TELEFAX:  302-773-0164
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  606 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-556-978B-21
```

```
  Query Match           11.4%;  Score 194.5;  DB 4;  Length 606;
  Best Local Similarity  26.4%;  Pred. No. 3.1e-09;
  Matches   96;  Conservative   31;  Mismatches  176;  Indels   61;  Gaps   12;

Qy      25 AASAASAATGNAAIGDVVKNSGAAAKGGEAASVNGIAKGIKGIVDAAGKADAKEGKLDAT 84
           | :||:|| | |  | :         |  ||:  |  :|  | : : |      | | :
Db      42 AGAAAAAAAGGAGQGGLGSQGAGQGAGAAAAAAGGAGQGGYGGLGSQGAGQGGYGGLGSQ 101

Qy      85 GA------EGTTNVNAGKLFVKRAADDGG----DADDAGKA-----AAAVAASAATGNAA 129
           ||       |     ||       |  ||     :  ||:       |  ||:|| | |
Db     102 GAGRGGQGAGAAAAAAG------GAGQGGYGGLGSQGAGRGGLGGQGAGAAAAAAAGGAG 155

Qy     130 IGDVVNGDVAKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGA-EGTTNADAGK 188
           | : :    :  |  ||:  |  :|  | : :        | | : ||  |   | |
Db     156 QGGLGSQGAGQGAGAAAAAAGGAGQGGYGGLGSQGAGQGGYGGLGSQGAGRGGQGAGAAA 215

Qy     189 LFVKNAGNVG-----------GEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKK 237
             ||  |             |  |  |  ||| ||  |          | : |   ||
Db     216 AAAGGAGQGGYGGLGSQGAGRGGLGGQGAGAAAAAAAGG---------AGQGGLGSQG-- 264

Qy     238 AADRTNPIDAAIGGAGDNDAAAAFATMKKDDQIAAAMVLRGMAKDGQFALKDAAAA---- 293
           |        || ||||         :           :  :|  : || |   ||||
Db     265 AGQGAGAAAAAAGGAGQG-GYGGLGSQGAGQGGYGGLGSQGAGRGGQGAGAAAAAAGGAG 323

Qy     294 --------HEGTVKNAVDIIKAAAEAASAASAATGSAAIGDVVNGNGATAKGGDAKSVNG 345
                  :|  :  :    | | ||:||  | |   :|    | ||   |  | :  |
Db     324 QGGYGGLGSQGAGRGGLGGQGAGAAAAAAAGGA-GQGGLGSQGAGQGA---GAAAAAGG 379

Qy     346 IAKG 349
             :|
Db     380 AGQG 383
```

```
RESULT  14
US-09-247-806-4
; Sequence 4, Application US/09247806
; Patent No. 6280747
; GENERAL INFORMATION:
;  APPLICANT: PHILLIPPE, Michel
;  APPLICANT: GARSON, Jean-Claude
;  APPLICANT: ARRAUDEAU, Jean-Pierre
;  TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT
;  TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN
;  TITLE OF INVENTION: ANALOG
;  FILE REFERENCE: 6388-0365-0
;  CURRENT APPLICATION NUMBER: US/09/247,806
;  CURRENT FILING DATE: 1999-02-11
;  EARLIER APPLICATION NUMBER: FR 98/01614
;  EARLIER FILING DATE: 1998-02-11
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 606
;   TYPE: PRT
;   ORGANISM: Nephila clavipes
US-09-247-806-4
```

```
  Query Match           11.4%;  Score 194.5;  DB 4;  Length 606;
```

```
  Best Local Similarity  26.4%;  Pred. No. 3.1e-09;
  Matches   96;  Conservative   31;  Mismatches  176;  Indels   61;  Gaps   12;

Qy      25 AASAASAATGNAAIGDVVKNSGAAAKGGEAASVNGIAKGIKGIVDAAGKADAKEGKLDAT 84
           | :||:|| | |  | :         |  ||:  |  :|  | : : |      | | :
Db      42 AGAAAAAAAGGAGQGGLGSQGAGQGAGAAAAAAGGAGQGGYGGLGSQGAGQGGYGGLGSQ 101

Qy      85 GA------EGTTNVNAGKLFVKRAADDGG----DADDAGKA-----AAAVAASAATGNAA 129
           ||       |     ||       |  ||     :  ||:       |  ||:|| | |
Db     102 GAGRGGQGAGAAAAAAG------GAGQGGYGGLGSQGAGRGGLGGQGAGAAAAAAAGGAG 155

Qy     130 IGDVVNGDVAKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGA-EGTTNADAGK 188
           | : :    :  |  ||:  |  :|  | : :        | | : ||  |   | |
Db     156 QGGLGSQGAGQGAGAAAAAAGGAGQGGYGGLGSQGAGQGGYGGLGSQGAGRGGQGAGAAA 215

Qy     189 LFVKNAGNVG-----------GEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKK 237
             ||  |             |  |  |  ||| ||  |          | : |   ||
Db     216 AAAGGAGQGGYGGLGSQGAGRGGLGGQGAGAAAAAAAGG---------AGQGGLGSQG-- 264

Qy     238 AADRTNPIDAAIGGAGDNDAAAAFATMKKDDQIAAAMVLRGMAKDGQFALKDAAAA---- 293
           |        || ||||         :           :  :|  : || |   ||||
Db     265 AGQGAGAAAAAAGGAGQG-GYGGLGSQGAGQGGYGGLGSQGAGRGGQGAGAAAAAAGGAG 323

Qy     294 --------HEGTVKNAVDIIKAAAEAASAASAATGSAAIGDVVNGNGATAKGGDAKSVNG 345
                  :|  :  :    | | ||:||  | |   :|    | ||   |  | :  |
Db     324 QGGYGGLGSQGAGRGGLGGQGAGAAAAAAAGGA-GQGGLGSQGAGQGA---GAAAAAGG 379

Qy     346 IAKG 349
             :|
Db     380 AGQG 383
```

```
RESULT  15
US-07-941-523-24
; Sequence 24, Application US/07941523
; Patent No. 5571718
;  GENERAL INFORMATION:
;    APPLICANT:  Dunn, John J
;    APPLICANT:  Barbour, Alan G
;    TITLE OF INVENTION:  Cloning and Expression of Borrelia
;    TITLE OF INVENTION:  Lipoproteins
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  01730
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/941,523
;      FILING DATE:  19920908
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Granahan, Patricia
;      REGISTRATION NUMBER:  32,227
;      REFERENCE/DOCKET NUMBER:  BNL90-01A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 861-6240
;      TELEFAX:  (617) 861-9540
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  344 amino acids
;      TYPE:  AMINO ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-941-523-24
```

Query Match 11.3%; Score 193; DB 1; Length 344;
Best Local Similarity 25.6%; Pred. No. 2.1e-09;
Matches 91; Conservative 50; Mismatches 149; Indels 66; Gaps 14;

```
Qy      18 MAKAAEEAASAASAATGNAAIGDVVKNSGAAAKGGEAASVNGIAKGIKGIVDAAGKADAK 77
           ||    ||      :   |  :|: :   | :|:    | :    : :  ::    |:|
Db       1 MAGQQPEAGKTGVSGGVNGNLGNSLMELGRSAENAFYAFI----ELVSDVLGFTAKSDTT 56

Qy      78 E----GKLDATGAE-GTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAA-------- 124
           :    |  :: ||: |  : :  :: ||  |: | |  |: |     | : |
Db      57 KQEVGGYFNSLGAKLGEASNDLEQVAVK--AETGVDKSDSSKNPIREAVNEAKEVLGTLK 114

Qy     125 ---TGNAAIGD---VVNGDVAKAKGGDAA--SVNGIAKGIKGIVDAAEKADAKEGKLNAA 176
                |||   |   : |   |  ||   :    | :: || ||  |  |  |: |
Db     115 GYVESLGTIGDSNPVGYANNAAGSGTTAADDELRKAFKALQEIVKAATDAGVKALKIGAT 174

Qy     177 GAE---GTTNADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEK 233
           :   |  | :  |:   : ||    | |  |||| :::|||::| :||        |
Db     175 TLQANGGADNKEGAKILATSGGNPA--AADVAKAAAILSSVSGEEMLSSIV--------K 224

Qy     234 QGKKAADRTNPIDAAIGGAGD-NDAAAAFATMKKDDQIAAAMVLRGMAKDGQFALKDAAA 292
           |:         || :  | | | :| :||    |  :| |   :  |  |  ||:
Db     225 SGEN--------DAQLAAAADGNTSAISFAKGGSDAHLAGANTPKAAAVAGGIALR---- 272

Qy     293 AHEGTVKNAVDIIKAAAEAASAASAATGSAAIGDVVNGNGATAKGGDAKSVNGIAK 348
                   ::|    || ||  |||    |  | | |  |     ::|  : |
Db     273 ----------SLVKTGKLAAGAADNATGG---GKEVQGVGVAAANKLLRAVEDVIK 315
```

Search completed: January 10, 2002, 14:03:56
Job time: 90 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 10, 2002, 14:02:26 ; Search time 184.82 Seconds
(without alignments)
524.306 Million cell updates/sec

Title: US-09-445-803-2
Perfect score: 1708
Sequence: 1 KNNDHDNHKGTVKNAVDMAK.....GNGATAKGGDAKSVNGIAKG 349

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 3148936 seqs, 277657034 residues

Total number of hits satisfying chosen parameters: 3148936

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_Main:*
1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US080_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US081_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US082_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US083_COMB.pep:*
8: /cgn2_6/ptodata/2/paa/US084_COMB.pep:*
9: /cgn2_6/ptodata/2/paa/US085_COMB.pep:*
10: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
11: /cgn2_6/ptodata/2/paa/US087_COMB.pep:*
12: /cgn2_6/ptodata/2/paa/US088_COMB.pep:*
13: /cgn2_6/ptodata/2/paa/US089_COMB.pep:*
14: /cgn2_6/ptodata/2/paa/US090_COMB.pep:*
15: /cgn2_6/ptodata/2/paa/US091_COMB.pep:*
16: /cgn2_6/ptodata/2/paa/US092_COMB.pep:*
17: /cgn2_6/ptodata/2/paa/US093_COMB.pep:*
18: /cgn2_6/ptodata/2/paa/US094_COMB.pep:*
19: /cgn2_6/ptodata/2/paa/US095_COMB.pep:*
20: /cgn2_6/ptodata/2/paa/US096_COMB.pep:*
21: /cgn2_6/ptodata/2/paa/US097_COMB.pep:*
22: /cgn2_6/ptodata/2/paa/US098_COMB.pep:*
23: /cgn2_6/ptodata/2/paa/US099_COMB.pep:*
24: /cgn2_6/ptodata/2/paa/US60_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1708	100.0	349	18	US-09-445-803-2	Sequence 2, Appli
2	1702	99.6	349	17	US-09-300-971A-9	Sequence 9, Appli
3	532	31.1	323	17	US-09-300-971A-11	Sequence 11, Appl
4	480	28.1	189	17	US-09-300-971A-10	Sequence 10, Appl
5	437.5	25.6	356	1	PCT-US97-02952-2	Sequence 2, Appli
6	437.5	25.6	356	15	US-09-125-619-2	Sequence 2, Appli
7	437.5	25.6	356	15	US-09-125-619-13	Sequence 13, Appl
8	417.5	24.4	208	15	US-09-125-619-47	Sequence 47, Appl
9	416	24.4	211	15	US-09-125-619-37	Sequence 37, Appl
10	412.5	24.2	214	15	US-09-125-619-34	Sequence 34, Appl
11	409.5	24.0	212	15	US-09-125-619-31	Sequence 31, Appl
12	409.5	24.0	212	15	US-09-125-619-43	Sequence 43, Appl
13	409.5	24.0	216	15	US-09-125-619-38	Sequence 38, Appl
14	408	23.9	215	15	US-09-125-619-36	Sequence 36, Appl
15	405.5	23.7	214	15	US-09-125-619-35	Sequence 35, Appl
16	405	23.7	209	15	US-09-125-619-33	Sequence 33, Appl
17	405	23.7	213	15	US-09-125-619-39	Sequence 39, Appl
18	404	23.7	211	15	US-09-125-619-40	Sequence 40, Appl
19	403	23.6	209	15	US-09-125-619-42	Sequence 42, Appl
20	398.5	23.3	216	15	US-09-125-619-46	Sequence 46, Appl
21	398	23.3	209	15	US-09-125-619-45	Sequence 45, Appl
22	397.5	23.3	212	15	US-09-125-619-48	Sequence 48, Appl
23	392.5	23.0	212	15	US-09-125-619-32	Sequence 32, Appl
24	390	22.8	189	15	US-09-125-619-20	Sequence 20, Appl
25	389.5	22.8	212	15	US-09-125-619-44	Sequence 44, Appl
26	387	22.7	187	15	US-09-125-619-23	Sequence 23, Appl
27	384	22.5	190	15	US-09-125-619-15	Sequence 15, Appl
28	382.5	22.4	212	15	US-09-125-619-41	Sequence 41, Appl
29	379	22.2	190	15	US-09-125-619-19	Sequence 19, Appl
30	377.5	22.1	195	15	US-09-125-619-17	Sequence 17, Appl
31	373	21.8	197	15	US-09-125-619-25	Sequence 25, Appl
32	364.5	21.3	169	15	US-09-125-619-22	Sequence 22, Appl
33	363.5	21.3	194	15	US-09-125-619-28	Sequence 28, Appl
34	362.5	21.2	169	15	US-09-125-619-24	Sequence 24, Appl
35	361.5	21.2	190	15	US-09-125-619-29	Sequence 29, Appl
36	359	21.0	179	15	US-09-125-619-18	Sequence 18, Appl
37	359	21.0	190	15	US-09-125-619-27	Sequence 27, Appl
38	345.5	20.2	189	15	US-09-125-619-26	Sequence 26, Appl
39	341	20.0	169	15	US-09-125-619-16	Sequence 16, Appl
40	319	18.7	158	15	US-09-125-619-21	Sequence 21, Appl
41	266.5	15.6	339	15	US-09-125-619-14	Sequence 14, Appl
42	242	14.2	157	15	US-09-125-619-30	Sequence 30, Appl
43	226.5	13.3	456	24	US-60-191-637-20600	Sequence 20600, A
44	226.5	13.3	456	24	US-60-191-681-16220	Sequence 16220, A
45	214	12.5	681	1	PCT-US00-30086-4	Sequence 4, Appli

ALIGNMENTS

```
RESULT 1
US-09-445-803-2
; Sequence 2, Application US/09445803
;   GENERAL INFORMATION:
;        APPLICANT: Adminis. of Tulane Educational, Fund
;                   Philipp, Mario T.
;        TITLE OF INVENTION: Surface Antigens and Proteins Useful in
;                            Compositions for the Diagnosis and Prevention of Lyme
;                            Disease
;        NUMBER OF SEQUENCES: 14
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Howson and Howson
;             STREET: Spring House Corporate Cntr., P.O. Box 457
;             CITY: Spring House
;             STATE: Pennsylvania
;             COUNTRY: USA
;             ZIP: 19477
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/445,803
;             FILING DATE: 13-Dec-1999
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US 60/051,271
;             FILING DATE: 30-JUN-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Bak, Mary E.
;             REGISTRATION NUMBER: 31,215
```

```
;             REFERENCE/DOCKET NUMBER: TUL2APCT
;         TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 215-540-9200
;             TELEFAX: 215-540-5818
;    INFORMATION FOR SEQ ID NO: 2:
;         SEQUENCE CHARACTERISTICS:
;             LENGTH: 349 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;         MOLECULE TYPE: protein
;         SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-445-803-2


  Query Match           100.0%;  Score 1708;  DB 18;  Length 349;
  Best Local Similarity  100.0%;  Pred. No. 9.7e-137;
  Matches  349;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 KNNDHDNHKGTVKNAVDMAKAAEEAASAASAATGNAAIGDVVKNSGAAAKGGEAASVNGI 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 KNNDHDNHKGTVKNAVDMAKAAEEAASAASAATGNAAIGDVVKNSGAAAKGGEAASVNGI 60

Qy      61 AKGIKGIVDAAGKADAKEGKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AKGIKGIVDAAGKADAKEGKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVA 120

Qy     121 ASAATGNAAIGDVVNGDVAKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGAEG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ASAATGNAAIGDVVNGDVAKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGAEG 180

Qy     181 TTNADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAAD 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TTNADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAAD 240

Qy     241 RTNPIDAAIGGAGDNDAAAAFATMKKDDQIAAAMVLRGMAKDGQFALKDAAAAHEGTVKN 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 RTNPIDAAIGGAGDNDAAAAFATMKKDDQIAAAMVLRGMAKDGQFALKDAAAAHEGTVKN 300

Qy     301 AVDIIKAAAEAASAASAATGSAAIGDVVNGNGATAKGGDAKSVNGIAKG 349
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AVDIIKAAAEAASAASAATGSAAIGDVVNGNGATAKGGDAKSVNGIAKG 349


RESULT   2
US-09-300-971A-9
; Sequence 9, Application US/09300971A
; GENERAL INFORMATION:
;  APPLICANT: Philipp, Mario T.
;  APPLICANT:  Liang, Fang Ting
;  TITLE OF INVENTION: Novel Peptides and Assays for the Diagnosis of Lyme Disease
;  FILE REFERENCE: TUL3USA
;  CURRENT APPLICATION NUMBER: US/09/300,971A
;  CURRENT FILING DATE:  1999-04-28
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
;   LENGTH: 349
;   TYPE: PRT
;   ORGANISM: Borrelia garinii
US-09-300-971A-9


  Query Match           99.6%;  Score 1702;  DB 17;  Length 349;
  Best Local Similarity  99.7%;  Pred. No. 3.2e-136;
  Matches  348;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 KNNDHDNHKGTVKNAVDMAKAAEEAASAASAATGNAAIGDVVKNSGAAAKGGEAASVNGI 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 KNNDHDNHKGTVKNAVDMAKAAEEAASAASAATGNAAIGDVVKNSGAAAKGGEAASVNGI 60

Qy      61 AKGIKGIVDAAGKADAKEGKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AKGIKGIVDAAGKADAKEGKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVA 120

Qy     121 ASAATGNAAIGDVVNGDVAKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGAEG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ASAATGNAAIGDVVNGDVAKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGAEG 180

Qy     181 TTNADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAAD 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TTNADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAAD 240

Qy     241 RTNPIDAAIGGAGDNDAAAAFATMKKDDQIAAAMVLRGMAKDGQFALKDAAAAHEGTVKN 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 ATNPIDAAIGGAGDNDAAAAFATMKKDDQIAAAMVLRGMAKDGQFALKDAAAAHEGTVKN 300

Qy     301 AVDIIKAAAEAASAASAATGSAAIGDVVNGNGATAKGGDAKSVNGIAKG 349
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AVDIIKAAAEAASAASAATGSAAIGDVVNGNGATAKGGDAKSVNGIAKG 349


RESULT   3
US-09-300-971A-11
; Sequence 11, Application US/09300971A
; GENERAL INFORMATION:
;  APPLICANT: Philipp, Mario T.
;  APPLICANT:  Liang, Fang Ting
;  TITLE OF INVENTION: Novel Peptides and Assays for the Diagnosis of Lyme Disease
;  FILE REFERENCE: TUL3USA
;  CURRENT APPLICATION NUMBER: US/09/300,971A
;  CURRENT FILING DATE:  1999-04-28
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
;   LENGTH: 323
;   TYPE: PRT
;   ORGANISM: Borrelia burgdorferi
US-09-300-971A-11


  Query Match           31.1%;  Score 532;  DB 17;  Length 323;
  Best Local Similarity  42.9%;  Pred. No. 9.5e-37;
  Matches  147;  Conservative   42;  Mismatches  110;  Indels   44;  Gaps   11;

Qy      31 AATGNAAIGDVVK--NSGAAAKGGEAASVNGIAKGIKGIVDAAGKADAKEGKLDATGAEG 88
           | :|   :  :||   :||| : ||    | ||    |  | || :| :|     | :
Db       1 AVSGEQILSAIVKAAAAGAADQDGEKP---GDAK--NPIAAAIGKGNADDGADFGDGMKK 55

Qy      89 TTNVNA----------GKLFVKRAADDGGDADDAGKAAAAV----------AASAATGNA 128
             : |           || ||:  |: | |: | | |: :           |  |::| |
Db      56 DDQIAAAIALRGMAKDGKFAVKK--DEKGKAEGAIKGASELLDKLVKAVKTAEGASSGTA 113

Qy     129 AIGDVVNGDVAKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGAEGTTNADAGK 188
           |||:||:     ||  |  || ||||||| ||:|| ::    ||  | |:|  |  |||
Db     114 AIGEVVDN---AAKAADKDSVTGIAKGIKEIVEAAGGSE----KLKVAAAKGENNKGAGK 166

Qy     189 LFVKNAGNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAADRTNPIDAA 248
           || |   |  |::  | ||| ||:|||||||| ||| || : |:::|||  :  ||| ||
Db     167 LFGKAGANAHGDSEAASKAAGAVSAVSGEQILSAIVKAAGEAGDQEGKKPEEAKNPIAAA 226

Qy     249 IGGAGDNDAAAAFAT--MKKDDQIAAAMVLRGMAKDGQFALKDAAAAHEGTVKNAVDIIK 306
           |   || |  | |    ||||||||||: ||||||||:||:||     || :|   :::
Db     227 I---GDKDGDAEFNQDGMKKDDQIAAAIALRGMAKDGKFAVKDGGEKAEGAIKGVSELLD 283

Qy     307 AAAEAASAASAATGSAAIGDVVNGNGATAKGGDAKSVNGIAKG 349
                :|  |::|:||||:||      ||  |  || |||||
Db     284 LVKAVKTAEGASSGTAAIGEVV---ADAAKVADKASVTGIAKG 323


RESULT   4
US-09-300-971A-10
; Sequence 10, Application US/09300971A
; GENERAL INFORMATION:
;  APPLICANT: Philipp, Mario T.
;  APPLICANT:  Liang, Fang Ting
;  TITLE OF INVENTION: Novel Peptides and Assays for the Diagnosis of Lyme Disease
```

```
;  FILE REFERENCE: TUL3USA
;  CURRENT APPLICATION NUMBER: US/09/300,971A
;  CURRENT FILING DATE:  1999-04-28
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
;   LENGTH: 189
;   TYPE: PRT
;   ORGANISM: Borrelia burgdorferi
US-09-300-971A-10


  Query Match             28.1%;  Score 480;  DB 17;  Length 189;
  Best Local Similarity   64.6%;  Pred. No. 1.2e-32;
  Matches  115;  Conservative   16;  Mismatches   39;  Indels    8;  Gaps    5;

Qy    117 AAVAASAATGN-AAIGDVV---NGDVAKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGK 172
          |  ||:||:|:   || ||   |   | ||||:  |||||| ||||||:|||||  ||||
Db      1 ATEAATAASGDKEMIGKVVKVTNAGAAAAKGGEEKSVNGIASGIKGIVEAAEKA-GKEGK 59

Qy    173 LNAAGAEGTTNADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGE 232
          | ||  :|  :| ||||| ||| | || |  | ||||||:||||:||||||| ||  | |
Db     60 LEAAAGDGNKDACAGKLFAKNAANGGGGAAAAEKAAAAVSAVSGKQILKAIVDAA--GKE 117

Qy    233 KQG-KKAADRTNPIDAAIGGAGDNDAAAAFATMKKDDQIAAAMVLRGMAKDGQFALKD 289
          ::|     : ||||:|||| |||||||    |||:||||||:|||||||||:|||||
Db    118 EKGVADVKEATNPIEAAIGSTGDNDAAAFQDEMKKNDQIAAAIVLRGMAKDGEFALKD 175


RESULT   5
PCT-US97-02952-2
; Sequence 2, Application PC/TUS9702952
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  VMP-Like Sequences of Pathogenic Borrelia
;    NUMBER OF SEQUENCES:  8
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US97/02952
;      FILING DATE:
;      CLASSIFICATION:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  356 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US97-02952-2


  Query Match             25.6%;  Score 437.5;  DB 1;  Length 356;
  Best Local Similarity   46.9%;  Pred. No. 1.2e-28;
  Matches  127;  Conservative   26;  Mismatches   87;  Indels   31;  Gaps   11;

Qy     79 GKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGDVVNGDV 138
          || |:||: ||         |: |  :  :  |    |   |  |::| ||||:||  |
Db    103 GKPDSTGSVGTA--------VEGAIKEVSELLDKLVKAVKTAEGASSGTAAIGEVV-ADA 153

Qy    139 AKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNA-AGAEGTTNADAGKLFVKNAGNV 197
            ||  | ||| |||||||  ||:||  ::    || | | |:|  |  |||||  |
Db    154 DAAKVADKASVKGIAKGIKEIVEAAGGSE----KLKAVAAAKGENNKGAGKLFGKAGAAA 209

Qy    198 GGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAADRTNPIDAAIGGAGDNDA 257
          |::  | ||| ||:|||||||| ||| || |  |: |||  :  ||| |||   || |
Db    210 HGDSEAASKAAGAVSAVSGEQILSAIVTAA-DAAEQDGKKPEEAKNPIAAAI---GDKDG 265

Qy    258 AAAFA--TMKKDDQIAAAMVLRGMAKDGQFALKDAAAAHEGTVKNAVDIIKAAAEAA--S 313
          | |    ||||||||||: ||||||||:||:||      |  : |   || |||:|
Db    266 GAEFGQDEMKKDDQIAAAIALRGMAKDGKFAVKD------GEKEKAEGAIKGAAESAVRK 319

Qy    314 AASAATGSAAIGDVVNGNGATAKGGDAKSVN 344
              | ||   ||| |: :|    |   |: :
Db    320 VLGAITG--LIGDAVS-SGLRKVGDSVKAAS 347


RESULT   6
US-09-125-619-2
; Sequence 2, Application US/09125619
; GENERAL INFORMATION:
;  APPLICANT:  NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION:  VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE:  UTSH:234
;  CURRENT APPLICATION NUMBER: US/09/125,619
;  CURRENT FILING DATE:  1999-01-27
;  NUMBER OF SEQ ID NOS:  48
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 356
;   TYPE: PRT
;   ORGANISM: Borrelia burgdorferi
US-09-125-619-2


  Query Match             25.6%;  Score 437.5;  DB 15;  Length 356;
  Best Local Similarity   46.9%;  Pred. No. 1.2e-28;
  Matches  127;  Conservative   26;  Mismatches   87;  Indels   31;  Gaps   11;

Qy     79 GKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGDVVNGDV 138
          || |:||: ||         |: |  :  :  |    |   |  |::| ||||:||  |
Db    103 GKPDSTGSVGTA--------VEGAIKEVSELLDKLVKAVKTAEGASSGTAAIGEVV-ADA 153

Qy    139 AKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNA-AGAEGTTNADAGKLFVKNAGNV 197
            ||  | ||| |||||||  ||:||  ::    || | | |:|  |  |||||  |
Db    154 DAAKVADKASVKGIAKGIKEIVEAAGGSE----KLKAVAAAKGENNKGAGKLFGKAGAAA 209

Qy    198 GGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAADRTNPIDAAIGGAGDNDA 257
          |::  | ||| ||:|||||||| ||| || |  |: |||  :  ||| |||   || |
Db    210 HGDSEAASKAAGAVSAVSGEQILSAIVTAA-DAAEQDGKKPEEAKNPIAAAI---GDKDG 265

Qy    258 AAAFA--TMKKDDQIAAAMVLRGMAKDGQFALKDAAAAHEGTVKNAVDIIKAAAEAA--S 313
          | |    ||||||||||: ||||||||:||:||      |  : |   || |||:|
Db    266 GAEFGQDEMKKDDQIAAAIALRGMAKDGKFAVKD------GEKEKAEGAIKGAAESAVRK 319

Qy    314 AASAATGSAAIGDVVNGNGATAKGGDAKSVN 344
              | ||   ||| |: :|    |   |: :
Db    320 VLGAITG--LIGDAVS-SGLRKVGDSVKAAS 347


RESULT   7
US-09-125-619-13
; Sequence 13, Application US/09125619
; GENERAL INFORMATION:
;  APPLICANT:  NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION:  VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE:  UTSH:234
;  CURRENT APPLICATION NUMBER: US/09/125,619
;  CURRENT FILING DATE:  1999-01-27
;  NUMBER OF SEQ ID NOS:  48
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 13
;   LENGTH: 356
;   TYPE: PRT
```

```
;   ORGANISM: Borrelia burgdorferi
US-09-125-619-13


  Query Match             25.6%;  Score 437.5;  DB 15;  Length 356;
  Best Local Similarity   46.9%;  Pred. No. 1.2e-28;
  Matches  127;  Conservative   26;  Mismatches   87;  Indels   31;  Gaps   11;

Qy      79 GKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGDVVNGDV 138
           || |:||: ||         |: |  :  :  |    |   | |::| ||||:||  |
Db     103 GKPDSTGSVGTA--------VEGAIKEVSELLDKLVKAVKTAEGASSGTAAIGEVV-ADA 153

Qy     139 AKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNA-AGAEGTTNADAGKLFVKNAGNV 197
             ||  | ||| ||||||| ||:||  ::    || | | |:|  |  ||||| |
Db     154 DAAKVADKASVKGIAKGIKEIVEAAGGSE----KLKAVAAAKGENNKGAGKLFGKAGAAA 209

Qy     198 GGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAADRTNPIDAAIGGAGDNDA 257
           |::  | ||| ||:|||||||| ||| || |  |: |||  :  ||| |||   || |
Db     210 HGDSEAASKAAGAVSAVSGEQILSAIVTAA-DAAEQDGKKPEEAKNPIAAAI---GDKDG 265

Qy     258 AAAFA--TMKKDDQIAAAMVLRGMAKDGQFALKDAAAAHEGTVKNAVDIIKAAAEAA--S 313
           | |    |||||||||||: ||||||||:||:||       |  : |   || |||:|
Db     266 GAEFGQDEMKKDDQIAAAIALRGMAKDGKFAVKD------GEKEKAEGAIKGAAESAVRK 319

Qy     314 AASAATGSAAIGDVVNGNGATAKGGDAKSVN 344
             | ||   ||| |: :|    |   |: :
Db     320 VLGAITG--LIGDAVS-SGLRKVGDSVKAAS 347


RESULT   8
US-09-125-619-47
; Sequence 47, Application US/09125619
; GENERAL INFORMATION:
;  APPLICANT:  NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION:  VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE:  UTSH:234
;  CURRENT APPLICATION NUMBER: US/09/125,619
;  CURRENT FILING DATE:  1999-01-27
;  NUMBER OF SEQ ID NOS:  48
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 47
;   LENGTH: 208
;   TYPE: PRT
;   ORGANISM: Borrelia burgdorferi
US-09-125-619-47


  Query Match             24.4%;  Score 417.5;  DB 15;  Length 208;
  Best Local Similarity   49.3%;  Pred. No. 2.9e-27;
  Matches  110;  Conservative   24;  Mismatches   68;  Indels   21;  Gaps    7;

Qy      79 GKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGDVVNGDV 138
           || |:||: ||         |: |  :  :  |    |   |  |::|  |||:||: |
Db       4 GKPDSTGSVGTA--------VEGAIKEVSELLDKLVTAVKTAEGASSGTDAIGEVVDND- 54

Qy     139 AKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGAEGTTNADAGKLFVKNAGNVG 198
             ||  | ||| ||||||| ||:||  ::    || || |:|  :  |||||  |   :
Db      55 --AKAADKASVTGIAKGIKEIVEAAGGSE----KLKAAAAKGENSKGAGKLFGKAGADAN 108

Qy     199 GEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAADRTNPIDAAIG-GAGDNDA 257
           |::  | ||| ||:|||||||| ||| || |  |: |:|  |  ||| |||| | || |
Db     109 GDSEAASKAAGAVSAVSGEQILSAIVTAA-DAAEQDGEKPGDAKNPIAAAIGKGDGDADF 167

Qy     258 AAAFATMKKDDQIAAAMVLRGMAKDGQFALK-DAAAAHEGTVK 299
                |||||||||: ||||||||:||:| |     || :|
Db     168 GDG---MKKDDQIAAAIALRGMAKDGKFAVKNDEKGKAEGAIK 207


RESULT   9
US-09-125-619-37
; Sequence 37, Application US/09125619
; GENERAL INFORMATION:
;  APPLICANT:  NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION:  VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE:  UTSH:234
;  CURRENT APPLICATION NUMBER: US/09/125,619
;  CURRENT FILING DATE:  1999-01-27
;  NUMBER OF SEQ ID NOS:  48
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 37
;   LENGTH: 211
;   TYPE: PRT
;   ORGANISM: Borrelia burgdorferi
US-09-125-619-37


  Query Match             24.4%;  Score 416;  DB 15;  Length 211;
  Best Local Similarity   49.1%;  Pred. No. 3.9e-27;
  Matches  109;  Conservative   26;  Mismatches   71;  Indels   16;  Gaps    6;

Qy      79 GKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGDVVNGDV 138
           || |:||: ||         |: |  :  | |    |   |  |::| ||||:||: :
Db       4 GKPDSTGSVGTA--------VEGAIKEVSGAADKLVKAVKTAEGASSGTAAIGEVVDNNN 55

Qy     139 AKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGAEGTTNADAGKLFVKNAGNVG 198
           | ||  | ||| ||||||| ||:||  ::    ||  | | | :|  ||||| |   :
Db      56 A-AKAADKASVTGIAKGIKEIVEAAGGSE----KLKVAAATGESNKGAGKLFGKAGADAH 110

Qy     199 GEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAADRTNPIDAAIGGAGDNDAA 258
           |::  | ||| ||:|||||||| ||| || |  :: |:|  |  ||| ||| | ||  |
Db     111 GDSEAASKAAGAVSAVSGEQILSAIVKAA-DAADQDGEKPGDAKNPIAAAI-GKGDGGAE 168

Qy     259 AAFATMKKDDQIAAAMVLRGMAKDGQFALK-DAAAAHEGTVK 299
             :   |||||||||: ||||||||:||:| :     || :|
Db     169 FSQDEMKKDDQIAAAIALRGMAKDGKFAVKNNEKGKAEGAIK 210


RESULT  10
US-09-125-619-34
; Sequence 34, Application US/09125619
; GENERAL INFORMATION:
;  APPLICANT:  NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION:  VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE:  UTSH:234
;  CURRENT APPLICATION NUMBER: US/09/125,619
;  CURRENT FILING DATE:  1999-01-27
;  NUMBER OF SEQ ID NOS:  48
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 34
;   LENGTH: 214
;   TYPE: PRT
;   ORGANISM: Borrelia burgdorferi
US-09-125-619-34


  Query Match             24.2%;  Score 412.5;  DB 15;  Length 214;
  Best Local Similarity   48.9%;  Pred. No. 7.9e-27;
  Matches  110;  Conservative   24;  Mismatches   72;  Indels   19;  Gaps    6;

Qy      79 GKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGDVVNGDV 138
           || |:||: ||         |: |  :  :  |    |   |  |::| ||||:||: |
```

```
Db      4 GKPDSTGSVGTA--------VEGAIKEVSELLDKLVKAVKTAEGASSGTAAIGEVVDND- 54

Qy    139 AKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGAEGTTNADAGKLFVKNAGNVG 198
            ||  | ||| |||||||  ||:||  ::    ||  | |:|  |  ||||| |
Db     55 --AKAADKASVTGIAKGIKEIVEAAGGSE----KLKVAAAKGENNKGAGKLFGKAGDAAN 108

Qy    199 GEAGDAGKAAAAVAAVSGEQILKAIVHAAKDG-GEKQGKKAADRTNPIDAAIG-GAGDND 256
          |::  | ||| ||:||||||| ||| ||  |  |: |:| |:  ||| |||| |  :|
Db    109 GDSEAASKAAGAVSAVSGEQILSAIVTAAAAGAAEQDGEKPAEAKNPIAAAIGKGNEENG 168

Qy    257 AAAAFATMKKDDQIAAAMVLRGMAKDGQFALK--DAAAAHEGTVK 299
          |      ||||||||||: |||||||||:||:|  |     || :|
Db    169 AEFNKEGMKKDDQIAAAIALRGMAKDGKFAVKKDDEKEKAEGAIK 213


RESULT  11
US-09-125-619-31
; Sequence 31, Application US/09125619
; GENERAL INFORMATION:
;  APPLICANT:  NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION:  VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE:  UTSH:234
;  CURRENT APPLICATION NUMBER: US/09/125,619
;  CURRENT FILING DATE:  1999-01-27
;  NUMBER OF SEQ ID NOS:  48
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 31
;   LENGTH: 212
;   TYPE: PRT
;   ORGANISM: Borrelia burgdorferi
US-09-125-619-31


  Query Match             24.0%;  Score 409.5;  DB 15;  Length 212;
  Best Local Similarity   49.8%;  Pred. No. 1.4e-26;
  Matches  112;  Conservative   21;  Mismatches   71;  Indels   21;  Gaps    8;

Qy     79 GKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGDVVNGDV 138
          || |:||: ||         |: |  :  :  |    |   |  |::| ||||:||  |
Db      4 GKPDSTGSVGTA--------VEGAIKEVSELLDKLVKAVKTAEGASSGTAAIGEVV-ADA 54

Qy    139 AKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNA-AGAEGTTNADAGKLFVKNAGNV 197
            ||  | ||| |||||||  ||:||  ::    || | |  |:|  |  ||||| |
Db     55 DAAKVADKASVKGIAKGIKEIVEAAGGSE----KLKAVAAAKGENNKGAGKLFGKAGAAA 110

Qy    198 GGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAADRTNPIDAAIGGAGDNDA 257
          |::  | ||| ||:||||||| ||| || |  |: |||  :  ||| |||   || |
Db    111 HGDSEAASKAAGAVSAVSGEQILSAIVTAA-DAAEQDGKKPEEAKNPIAAAI---GDKDG 166

Qy    258 AAAFA--TMKKDDQIAAAMVLRGMAKDGQFALKDAAAAH-EGTVK 299
          | |    ||||||||||: |||||||:||:||      || :|
Db    167 GAEFGQDEMKKDDQIAAAIALRGMAKDGKFAVKDGEKEKAEGAIK 211


RESULT  12
US-09-125-619-43
; Sequence 43, Application US/09125619
; GENERAL INFORMATION:
;  APPLICANT:  NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION:  VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE:  UTSH:234
;  CURRENT APPLICATION NUMBER: US/09/125,619
;  CURRENT FILING DATE:  1999-01-27
;  NUMBER OF SEQ ID NOS:  48
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 43
;   LENGTH: 212
;   TYPE: PRT
;   ORGANISM: Borrelia burgdorferi
US-09-125-619-43


  Query Match             24.0%;  Score 409.5;  DB 15;  Length 212;
  Best Local Similarity   49.8%;  Pred. No. 1.4e-26;
  Matches  112;  Conservative   21;  Mismatches   71;  Indels   21;  Gaps    8;

Qy     79 GKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGDVVNGDV 138
          || |:||: ||         |: |  :  :  |    |   |  |::| ||||:||  |
Db      4 GKPDSTGSVGTA--------VEGAIKEVSELLDKLVKAVKTAEGASSGTAAIGEVV-ADA 54

Qy    139 AKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNA-AGAEGTTNADAGKLFVKNAGNV 197
            ||  | ||| |||||||  ||:||  ::    || | |  |:|  |  ||||| |
Db     55 DAAKVADKASVKGIAKGIKEIVEAAGGSE----KLKAVAAAKGENNKGAGKLFGKAGAAA 110

Qy    198 GGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAADRTNPIDAAIGGAGDNDA 257
          |::  | ||| ||:||||||| ||| || |  |: |||  :  ||| |||   || |
Db    111 HGDSEAASKAAGAVSAVSGEQILSAIVTAA-DAAEQDGKKPEEAKNPIAAAI---GDKDG 166

Qy    258 AAAFA--TMKKDDQIAAAMVLRGMAKDGQFALKDAAAAH-EGTVK 299
          | |    ||||||||||: |||||||:||:||      || :|
Db    167 GAEFGQDEMKKDDQIAAAIALRGMAKDGKFAVKDGEKEKAEGAIK 211


RESULT  13
US-09-125-619-38
; Sequence 38, Application US/09125619
; GENERAL INFORMATION:
;  APPLICANT:  NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION:  VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE:  UTSH:234
;  CURRENT APPLICATION NUMBER: US/09/125,619
;  CURRENT FILING DATE:  1999-01-27
;  NUMBER OF SEQ ID NOS:  48
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 38
;   LENGTH: 216
;   TYPE: PRT
;   ORGANISM: Borrelia burgdorferi
US-09-125-619-38


  Query Match             24.0%;  Score 409.5;  DB 15;  Length 216;
  Best Local Similarity   48.7%;  Pred. No. 1.4e-26;
  Matches  111;  Conservative   26;  Mismatches   68;  Indels   23;  Gaps    8;

Qy     79 GKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGDVVNGDV 138
          || |:||: ||         |: |  :  :  |    |   |  |::|  |||:||: :
Db      4 GKPDSTGSVGTA--------VEGAIKEVSELLDKLVKAVKTAEEASSGTDAIGEVVD-NA 54

Qy    139 AKAKGGDAASVNGIAKGIKGIVDAA---EKADAKEGKLNAAGAEGTTNADAGKLFVK-NA 194
          | ||  | ||| |||||||  ||:||   ||  |  .   || | |  | :|||||  | :|
Db     55 AAAKAADKASVTGIAKGIKEIVEAAGGSEKLKA------AAAATGENNKEAGKLFGKVDA 108

Qy    195 GNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKD-GGEKQGKKAADRTNPIDAAIGGAG 253
          ||  |::  | ||| ||:||||||| ||| || :  |:::|||  :  ||| ||||
Db    109 GN-AGDSEAASKAAGAVSAVSGEQILSAIVKAAGEAAGDQEGKKPGEAKNPIAAAIGKGN 167

Qy    254 DNDAAAAFATMKKDDQIAAAMVLRGMAKDGQFALKDAAAAH--EGTVK 299
           :| |     ||||||||||: |||||||:||:|         || :|
Db    168 ADDGAEFGDGMKKDDQIAAAIALRGMAKDGKFAVKSGGEKEKAEGAIK 215
```

```
RESULT  14
US-09-125-619-36
; Sequence 36, Application US/09125619
; GENERAL INFORMATION:
;  APPLICANT:  NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION:  VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE:  UTSH:234
;  CURRENT APPLICATION NUMBER: US/09/125,619
;  CURRENT FILING DATE:  1999-01-27
;  NUMBER OF SEQ ID NOS:  48
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 36
;  LENGTH: 215
;  TYPE: PRT
;  ORGANISM: Borrelia burgdorferi
US-09-125-619-36


  Query Match             23.9%;  Score 408;  DB 15;  Length 215;
  Best Local Similarity   49.1%;  Pred. No. 1.9e-26;
  Matches  111;  Conservative   22;  Mismatches   73;  Indels   20;  Gaps    8;

Qy      79 GKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGDVVNGDV 138
           || |:||: ||         |: |  :  :  |    |   |  |::|  |||:|||: |
Db       4 GKPDSTGSVGTA--------VEGAIKEVSELLDKLVTAVKTAEGASSGTDAIGEVVDNDA 55

Qy     139 AKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGAEGTTNADAGKLFVKNAGNVG 198
           | ||  |  || ||||||| ||:||  ::    || |  |    |  ||||| | ||
Db      56 A-AKAADKESVTGIAKGIKEIVEAARGSE----KLKAVAAAKEGNEKAGKLFGK-AGAAA 109

Qy     199 GEAGDAGKAAAAVAAVSGEQILKAIVH-AAKDGGEKQGKKAADRTNPIDAAIGGAGDNDA 257
           |::  | ||| ||:|||||||| |||  ||   |:::|||  |  ||| ||| | || |
Db     110 GDSEAASKAAGAVSAVSGEQILSAIVKAAAAAAGDQEGKKPGDAKNPIAAAI-GKGDADD 168

Qy     258 AAAF-ATMKKDDQIAAAMVLRGMAKDGQFALKDAAAAH---EGTVK 299
            | |   ||||||||||: ||||||||:||:|         || :|
Db     169 GAEFDHEMKKDDQIAAAIALRGMAKDGKFAVKSGGGEKEKAEGAIK 214


RESULT  15
US-09-125-619-35
; Sequence 35, Application US/09125619
; GENERAL INFORMATION:
;  APPLICANT:  NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION:  VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE:  UTSH:234
;  CURRENT APPLICATION NUMBER: US/09/125,619
;  CURRENT FILING DATE:  1999-01-27
;  NUMBER OF SEQ ID NOS:  48
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 35
;  LENGTH: 214
;  TYPE: PRT
;  ORGANISM: Borrelia burgdorferi
US-09-125-619-35


  Query Match             23.7%;  Score 405.5;  DB 15;  Length 214;
  Best Local Similarity   49.1%;  Pred. No. 3.1e-26;
  Matches  112;  Conservative   21;  Mismatches   70;  Indels   25;  Gaps    8;

Qy      79 GKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGDVVNGDV 138
```

```
           || |:||: ||         |: |  :  :  |    |   |  |::|  |||:||  |
Db       4 GKPDSTGSVGTA--------VEGAIKEVSELLDKLVKAVKTAEGASSGTDAIGEVVADDN 55

Qy     139 AKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGAEGTTNADAGKLFVKNAGNVG 198
           | ||  | ||| |||||||| ||:||  ::    ||  | | | :|  |||||    | ||
Db      56 A-AKVADKASVTGIAKGIKEIVEAAGGSE----KLKVAAATGESNKGAGKLF----GKVG 106

Qy     199 GEAGD---AGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAADRTNPIDAAIG-GAGD 254
             |||   | ||| ||:||||||||| ||| ||    |::|||  :  ||| |||| |  :
Db     107 AHAGDSEAASKAAGAVSAVSGEQILSAIVKAA-GAAEQEGKKPEEAKNPIAAAIGKGDAE 165

Qy     255 NDAAAAFATMKKDDQIAAAMVLRGMAKDGQFALKDAAAAH---EGTVK 299
           | |       ||||||||||: ||||||||:||:|         || :|
Db     166 NGAEFNHDGMKKDDQIAAAIALRGMAKDGKFAVKSGGGEKGKAEGAIK 213


Search completed: January 10, 2002, 14:08:25
Job time: 359 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 10, 2002, 14:03:31 ; Search time 18.38 Seconds
(without alignments)
641.811 Million cell updates/sec

Title: US-09-445-803-2
Perfect score: 1708
Sequence: 1 KNNDHDNHKGTVKNAVDMAK.....GNGATAKGGDAKSVNGIAKG 349

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 155098 seqs, 33800819 residues

T[redacted] number of hits satisfying chosen parameters: 155098

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_New:*
1: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep:*
2: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep:*
3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep:*
4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep:*
5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep:*
6: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:*
7: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	220	12.9	1463	5	US-09-971-536-69	Sequence 69, Appl
2	207.5	12.1	606	5	US-09-861-597-6	Sequence 6, Appli
3	203.5	11.9	606	5	US-09-861-597-8	Sequence 8, Appli
4	197	11.5	651	5	US-09-861-597-1	Sequence 1, Appli
5	194.5	11.4	606	5	US-09-861-597-4	Sequence 4, Appli
6	182	10.7	1879	5	US-09-971-536-70	Sequence 70, Appl
7	172.5	10.1	410	5	US-09-500-746-17	Sequence 17, Appl
8	171.5	10.0	641	5	US-09-249-585A-3	Sequence 3, Appli
9	171	10.0	605	5	US-09-976-594-616	Sequence 616, App
10	167.5	9.8	624	5	US-09-952-267-7	Sequence 7, Appli
11	167.5	9.8	624	5	US-09-952-267A-7	Sequence 7, Appli
12	163.5	9.6	588	1	PCT-US01-20545-4	Sequence 4, Appli
13	158.5	9.3	529	5	US-09-861-597-2	Sequence 2, Appli
14	157.5	9.2	1531	5	US-09-995-493-44	Sequence 44, Appl
15	155	9.1	389	5	US-09-646-673A-172	Sequence 172, App
16	153.5	9.0	344	6	US-10-010-390-11	Sequence 11, Appl
17	151.5	8.9	731	5	US-09-340-736A-1	Sequence 1, Appli
18	151.5	8.9	731	5	US-09-340-736E-1	Sequence 1, Appli
19	149.5	8.8	914	5	US-09-975-143-47	Sequence 47, Appl
20	147.5	8.6	602	5	US-09-708-427-4646	Sequence 4646, Ap
21	147.5	8.6	609	5	US-09-708-427-4645	Sequence 4645, Ap
22	147.5	8.6	635	5	US-09-708-427-4644	Sequence 4644, Ap
23	147.5	8.6	1741	5	US-09-971-536-68	Sequence 68, Appl
24	145	8.5	889	5	US-09-952-267-15	Sequence 15, Appl
25	145	8.5	889	5	US-09-952-267A-15	Sequence 15, Appl
26	145	8.5	892	5	US-09-952-267-5	Sequence 5, Appli
27	145	8.5	892	5	US-09-952-267A-5	Sequence 5, Appli
28	145	8.5	2478	5	US-09-815-242-5816	Sequence 5816, Ap
29	145	8.5	2478	5	US-09-815-242-12967	Sequence 12967, A
30	144.5	8.5	730	5	US-09-708-427-1373	Sequence 1373, Ap
31	144.5	8.5	752	5	US-09-708-427-1372	Sequence 1372, Ap
32	140.5	8.2	6281	5	US-09-815-242-12996	Sequence 12996, A
33	140	8.2	1502	5	US-09-897-516-7005	Sequence 7005, Ap
34	139	8.1	452	5	US-09-708-427-11370	Sequence 11370, A
35	139	8.1	468	5	US-09-620-394B-2384	Sequence 2384, Ap
36	139	8.1	480	5	US-09-708-427-11369	Sequence 11369, A
37	139	8.1	485	5	US-09-620-394B-2383	Sequence 2383, Ap
38	139	8.1	490	5	US-09-708-427-11368	Sequence 11368, A
39	138.5	8.1	885	5	US-09-857-128-13	Sequence 13, Appl
40	138.5	8.1	928	5	US-09-857-128-12	Sequence 12, Appl
41	138.5	8.1	1503	5	US-09-708-427-23543	Sequence 23543, A
42	138	8.1	2368	5	US-09-815-242-5635	Sequence 5635, Ap
43	138	8.1	2368	5	US-09-815-242-12389	Sequence 12389, A
44	130.5	7.6	2086	5	US-09-815-242-5639	Sequence 5639, Ap
45	130.5	7.6	5795	5	US-09-815-242-12610	Sequence 12610, A

ALIGNMENTS

```
RESULT 1
US-09-971-536-69
; Sequence 69, Application US/09971536
; GENERAL INFORMATION:
;  APPLICANT: Glenn, Matthew
;  APPLICANT:  Havukkala, Ilkka
;  APPLICANT:  Bloksberg, Leonard
;  APPLICANT:  Lubbers, Mark
;  APPLICANT:  Dekker, James
;  APPLICANT:  Christensson, Anna
;  APPLICANT:  Holland, Ross
;  APPLICANT:  O'Toole, Paul
;  APPLICANT:  Reid, Julian
;  APPLICANT:  Coolbear, Timothy
;  TITLE OF INVENTION: Lactobacillus rhamnosus Polynucleotides, Polypeptides and Meth
;  TITLE OF INVENTION:  Using Them
;  FILE REFERENCE: 1043c2
;  CURRENT APPLICATION NUMBER: US/09/971,536
;  CURRENT FILING DATE:  2001-10-02
;  PRIOR APPLICATION NUMBER: U.S. No. 09/634,238
;  PRIOR FILING DATE: 2000-08-08
;  PRIOR APPLICATION NUMBER: U.S. No. 09/724,623
;  PRIOR FILING DATE: 2000-11-28
;  PRIOR APPLICATION NUMBER: PCT/NZ01/00160
;  PRIOR FILING DATE: 2001-08-08
;  NUMBER OF SEQ ID NOS: 83
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 69
;   LENGTH: 1463
;   TYPE: PRT
;   ORGANISM: Lactobacillus rhamnosus
US-09-971-536-69


  Query Match             12.9%;  Score 220;  DB 5;  Length 1463;
  Best Local Similarity   27.1%;  Pred. No. 2.8e-05;
  Matches   92;  Conservative   41;  Mismatches  145;  Indels   62;  Gaps    8;

Qy      21 AAEEAASAASAATGNAAIGDVVKNSGAAAKGGEAASVNGIAKGIKGIVDAAGKADAKEGK 80
           ||  ||:| |||   | :   |  | |||  |  ||           | :| ::|:|  :
Db     845 AASSAAAAMSAALSTAQVAAKVAVSDAAAAGSAAA-----------VASAAQSDSKNKQ 892

Qy      81 LDATGA--EGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGDVVNGDV 138
             |  |  :   ::|  |     |:     | :||:|: | :| |:  :::  :   |
Db     893 ATAATARSQALDDLNKIKSLTDYASGASSSASEAGQASTATSAYASAASSSASEA--GSY 950

Qy     139 AKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGAEGTTNADAGKLFVKNAGNVG 198
           |   |  |:             ||        |:  :|    :| | |   : |::|
Db     951 AHQAGSSAS------------DAV-------GQSGSAAQHASTAASAASSYPKDSGIQS 990
```

```
Qy    199 GEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAADRTNPIDAAIGGAGDNDAA 258
          :  | :|| | :  |      |:  :|     :| | ||       :|   |  | :|
Db    991 LASQAASEAAKASSNASAATSAAAVGFSAASDASEQAKTAASADVVASSAASTANSNASA 1050

Qy    259 AAFATMKKDDQIAAAMVLRGMAKDGQFALKDAAAAHEGTVKNAVDIIKAAAEAA------ 312
          || ||   | : ||                  ||:|   : | |  :   || ||
Db   1051 AASATKAGDSKAAA-------------GFSSAASAAASSAKGAEAVASEAASAAASDDSV 1097

Qy    313 --SAASAATG-----SAAIGDVVNGNGATAKGGDAKSVNG 345
          ||||||  |     ||| |   :   | |    |:   |
Db   1098 ASSAASAAAGFDKAASAAEGAASSAASAAASSAAAQGTRG 1137


RESULT   2
US-09-861-597-6
; Sequence 6, Application US/09861597
; GENERAL INFORMATION:
; APPLICANT: PHILLIPPE, Michel
; APPLICANT:  GARSON, Jean-Claude
; APPLICANT:  ARRAUDEAU, Jean-Pierre
;  TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT
;  TITLE OF INVENTION:  LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN
;  TITLE OF INVENTION:  ANALOG
;  FILE REFERENCE: 6388-0365-0
;  CURRENT APPLICATION NUMBER: US/09/861,597
;  CURRENT FILING DATE:  2001-05-22
;  PRIOR APPLICATION NUMBER: 09/247,806
;  PRIOR FILING DATE: 1999-02-11
;  PRIOR APPLICATION NUMBER: FR 98/01614
;  PRIOR FILING DATE: 1998-02-11
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 606
;   TYPE: PRT
;   ORGANISM: Nephila clavipes
US-09-861-597-6


  Query Match             12.1%;  Score 207.5;  DB 5;  Length 606;
  Best Local Similarity   26.8%;  Pred. No. 5.8e-05;
  Matches  102;  Conservative   32;  Mismatches  182;  Indels   65;  Gaps    9;

Qy     25 AASAASAATGNAAIGDVVKNSGAAAKGGEAASVNGIAKGIKGIV---------DAAGKAD 75
          | :||:|| | |  | :          |  ||:  |  :|  | :           || |
Db    229 AGAAAAAAAGGAGQGGLGSQGAGQGAGAAAAAAGGAGQGGYGGLGSQGAGRGGQGAGAAA 288

Qy     76 AKEGKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGDVVN 135
          |  |     |  |  :  ||:          |       |   || ||:|| | |  | : :
Db    289 AAAGGAGQGGYGGLGSQGAGQGGYGGLGSQGAGRGGLGGQGAAAAAAAAGGAGQGGLGS 348

Qy    136 GDVAKAKGGDAASVNGIAKGIKGIVD--------------AAEKADAKEGKLNAAGAEGT 181
             :  |  ||:  |  :|  | :                 ||     | :|     |::|
Db    349 QGAGQGAGAAAAAAGGAGQGGYGGLGSQGAGRGGQGAGAAAAAAGGAGQGGYGGLGSQGA 408

Qy    182 TNADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQIL-------KAIVHAAKDGGEKQ 234
             | |  : ||  |      || |||| |  :|:  |          |  ||  ||  |
Db    409 GQGGYGGLGSQGAGRGGLGGQGAGAAAAAAAGGAGQGGLGSQGAGQGAGAAAAAAGGAGQ 468

Qy    235 ------GKKAADRTNPIDAAIGGAGDNDAAAAFATMKK------DDQIAAAMVLRGMAKD 282
                | : | |        || |   ||||      :         | |       |:
Db    469 GGYGGLGSQGAGR--------GGQGAGAAAAAAGGAGQGGYGGLGSQGAGQGGYGGLGSQ 520

Qy    283 --------GQFALKDAAAAHEGTVKNAVDIIKAAAEAASAASAATGSA------AIGDVV 328
                  || |   |||| |  :  :   : | : | ||:|| | |        :|
Db    521 GAGRGGLGGQGAGAAAAAAGGAGQGGLG-SQGAGQGAGAAAAAAGGAGQGGYGGLGSQG 579

Qy    329 NGNGATAKGGDAKSVNGIAKG 349
             | |     |  | :  |  :|
Db    580 AGRGGQGAGAAAAAAGGAGQG 600


RESULT   3
US-09-861-597-8
; Sequence 8, Application US/09861597
; GENERAL INFORMATION:
;  APPLICANT: PHILLIPPE, Michel
;  APPLICANT:  GARSON, Jean-Claude
;  APPLICANT:  ARRAUDEAU, Jean-Pierre
;  TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT
;  TITLE OF INVENTION:  LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN
;  TITLE OF INVENTION:  ANALOG
;  FILE REFERENCE: 6388-0365-0
;  CURRENT APPLICATION NUMBER: US/09/861,597
;  CURRENT FILING DATE:  2001-05-22
;  PRIOR APPLICATION NUMBER: 09/247,806
;  PRIOR FILING DATE: 1999-02-11
;  PRIOR APPLICATION NUMBER: FR 98/01614
;  PRIOR FILING DATE: 1998-02-11
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
;   LENGTH: 606
;   TYPE: PRT
;   ORGANISM: Nephila clavipes
US-09-861-597-8


  Query Match             11.9%;  Score 203.5;  DB 5;  Length 606;
  Best Local Similarity   26.5%;  Pred. No. 9.7e-05;
  Matches  101;  Conservative   32;  Mismatches  183;  Indels   65;  Gaps    9;

Qy     25 AASAASAATGNAAIGDVVKNSGAAAKGGEAASVNGIAKGIKGIV---------DAAGKAD 75
          | :||:|| | |  | :          |  ||:  |  :|  | :           || |
Db     25 AGAAAAAAAGGAGQGGLGSQGAGQGAGAAAAAAGGAGQGGYGGLGSQGAGRGGQGAGAAA 84

Qy     76 AKEGKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGDVVN 135
          |  |     |  |  :  ||:          |       |   |  ||:|| | |  | : :
Db     85 AAAGGAGQGGYGGLGSQGAGQGGYGGLGSQGAGRGGLGGQGAGAAAAAAAGGAGQGGLGS 144

Qy    136 GDVAKAKGGDAASVNGIAKGIKGIVD--------------AAEKADAKEGKLNAAGAEGT 181
             :  |  ||:  |  :|  | :                 ||     | :|     |::|
Db    145 QGAGQGAGAAAAAAGGAGQGGYGGLGSQGAGRGGQGAGAAAAAAGGAGQGGYGGLGSQGA 204

Qy    182 TNADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQIL-------KAIVHAAKDGGEKQ 234
             | |  : ||  |      || |||| |  :|:  |          |  ||  ||  |
Db    205 GQGGYGGLGSQGAGRGGLGGQGAGAAAAAAAGGACQGGLGSQGAGQGAGAAAAAAGGAGQ 264

Qy    235 ------GKKAADRTNPIDAAIGGAGDNDAAAAFATMKK------DDQIAAAMVLRGMAKD 282
                | : | |        || |   ||||      :         | |       |:
Db    265 GGYGGLGSQGAGR--------GGQGAGAAAAAAGGAGQGGYGGLGSQGAGQGGYGGLGSQ 316

Qy    283 --------GQFALKDAAAAHEGTVKNAVDIIKAAAEAASAASAATGSA------AIGDVV 328
                  || |   |||| |  :  :   : | : | ||:|| | |        :|
Db    317 GAGRGGLGGQGAGAAAAAAGGAGQGGLG-SQGAGQGAGAAAAAAGGAGQGGYGGLGSQG 375

Qy    329 NGNGATAKGGDAKSVNGIAKG 349
             | |     |  | :  |  :|
Db    376 AGRGGQGAGAAAAAAGGAGQG 396


RESULT   4
US-09-861-597-1
; Sequence 1, Application US/09861597
; GENERAL INFORMATION:
;  APPLICANT: PHILLIPPE, Michel
;  APPLICANT:  GARSON, Jean-Claude
;  APPLICANT:  ARRAUDEAU, Jean-Pierre
;  TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT
;  TITLE OF INVENTION:  LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN
;  TITLE OF INVENTION:  ANALOG
;  FILE REFERENCE: 6388-0365-0
;  CURRENT APPLICATION NUMBER: US/09/861,597
;  CURRENT FILING DATE:  2001-05-22
;  PRIOR APPLICATION NUMBER: 09/247,806
```

```
;  PRIOR FILING DATE: 1999-02-11
;  PRIOR APPLICATION NUMBER: FR 98/01614
;  PRIOR FILING DATE: 1998-02-11
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 651
;   TYPE: PRT
;   ORGANISM: Nephila clavipes
US-09-861-597-1


  Query Match             11.5%;  Score 197;  DB 5;  Length 651;
  Best Local Similarity   26.1%;  Pred. No. 0.00024;
  Matches   98;  Conservative   32;  Mismatches  164;  Indels   82;  Gaps   13;

Qy      20 KAAEEAASAASAATGNAAIGDV-VKNSGAAAKGGE-----AASVNGIAKGIKGIVDAAGK 73
           : | : | |::|| | |  |       |  | :|||     ||:  |  :|  | :  |
Db     234 QGAGQGAGASAAAAGGAGQGGYGGLGSQGAGRGGEGAGAAAAAAGGAGQGGYGGLGGQGA 293

Qy      74 ADAKEGKLDATGA--EGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASA-------- 123
             | | : ||   |     ||       |   |     ||: | | ||:|
Db     294 GQGGYGGLGSQGAGRGGLGGQGAGAAAAGGAGQGGLGGQGAGQGAGAAAAAAGGAGQGGY 353

Qy     124 ------ATGNAAIGDVVNGDVAKAKGGDAA-------SVNGIAKGIKGIVDAAEKADAKE 170
                 |   :|    | || |  | |           |  :| :|    |  | |
Db     354 GGLGSQGAGRGGLGGQGAGAVAAAAAGGAGQGGYGGLGSQGAGRGGQG----AGAAAAAA 409

Qy     171 GKLNAAGAEGTTNADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDG 230
           |     |  |  |  ||:      | :||:   || |||| |  :|:     :      |
Db     410 GGAGQRGYGGLGNQGAGR------GGLGGQG--AGAAAAAAAGGAGQGGYGGL---GNQG 458

Qy     231 GEKQGKKAADRTNPIDAAIGGAGDN-----------------DAAAAFATMKKDDQIAAA 273
             : |: ||       || ||||                    ||||  |     : |
Db     459 AGRGGQGAA-------AAAGGAGQGGYGGLGSQGAGRGGQGAGAAAAAAVGAGQEGIRG- 510

Qy     274 MVLRGMAKDGQFALKDAAAAHEGTVKNAVDIIKAAAEAASAASAATGSAAIGDVVNGNGA 333
             :|  : |   |    :   |             : | ||:|| | |  | :
Db     511 ---QGAGQGGYGGLGSQGSGRGG----------LGGQGAGAAAAAAGGAGQGGLGGQGAG 557

Qy     334 TAKGGDAKSVNGIAKG 349
           |  | :  |: :|
Db     558 QGAGAAAAAAGGVRQG 573


RESULT   5
US-09-861-597-4
; Sequence 4, Application US/09861597
; GENERAL INFORMATION:
; APPLICANT: PHILLIPPE, Michel
; APPLICANT:  GARSON, Jean-Claude
;  APPLICANT:  ARRAUDEAU, Jean-Pierre
;  TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT
;  TITLE OF INVENTION:  LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN
;  TITLE OF INVENTION:  ANALOG
;  FILE REFERENCE: 6388-0365-0
;  CURRENT APPLICATION NUMBER: US/09/861,597
;  CURRENT FILING DATE:  2001-05-22
;  PRIOR APPLICATION NUMBER: 09/247,806
;  PRIOR FILING DATE: 1999-02-11
;  PRIOR APPLICATION NUMBER: FR 98/01614
;  PRIOR FILING DATE: 1998-02-11
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 606
;   TYPE: PRT
;   ORGANISM: Nephila clavipes
US-09-861-597-4


  Query Match             11.4%;  Score 194.5;  DB 5;  Length 606;
  Best Local Similarity   26.4%;  Pred. No. 0.00031;
  Matches   96;  Conservative   31;  Mismatches  176;  Indels   61;  Gaps   12;

Qy      25 AASAASAATGNAAIGDVVKNSGAAAKGGEAASVNGIAKGIKGIVDAAGKADAKEGKLDAT 84
           | :||:|| | |  | :          |  ||:  |  :|  | : : |       | | :
Db      42 AGAAAAAAAGGAGQGGLGSQGAGQGAGAAAAAAGGAGQGGYGGLGSQGAGQGGYGGLGSQ 101

Qy      85 GA------EGTTNVNAGKLFVKRAADDGG----DADDAGKA-----AAAVAASAATGNAA 129
           ||       |     ||       |  ||      :  ||:       |  ||:|| | |
Db     102 GAGRGGQGAGAAAAAAG------GAGQGGYGGLGSQGAGRGGLGGQGAGAAAAAAAGGAG 155

Qy     130 IGDVVNGDVAKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGA-EGTTNADAGK 188
           | : :    :  |  ||:  |  :|  | : :         | | : ||  |   | |
Db     156 QGGLGSQGAGQGAGAAAAAAGGAGQGGYGGLGSQGAGQGGYGGLGSQGAGRGGQGAGAAA 215

Qy     189 LFVKNAGNVG-----------GEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKK 237
             ||  |           |  |  |  ||| ||  |         | : |   ||
Db     216 AAAGGAGQGGYGGLGSQGAGRGGLGGQGAGAAAAAAAGG---------AGQGGLGSQG-- 264

Qy     238 AADRTNPIDAAIGGAGDNDAAAAFATMKKDDQIAAAMVLRGMAKDGQFALKDAAAA---- 293
           |         || ||||          :          :  :|  : || |   ||||
Db     265 AGQGAGAAAAAAGGAGQG-GYGGLGSQGAGQGGYGGLGSQGAGRGGQGAGAAAAAAGGAG 323

Qy     294 --------HEGTVKNAVDIIKAAAEAASAASAATGSAAIGDVVNGNGATAKGGDAKSVNG 345
                   :|  :  :    | | ||:||  | |   :|    | ||  |  | :  |
Db     324 QGGYGGLGSQGAGRGGLGGQGAGAAAAAAAGGA-GQGGLGSQGAGQGA---GAAAAAGG 379

Qy     346 IAKG 349
             :|
Db     380 AGQG 383


RESULT   6
US-09-971-536-70
; Sequence 70, Application US/09971536
; GENERAL INFORMATION:
;  APPLICANT: Glenn, Matthew
;  APPLICANT:  Havukkala, Ilkka
;  APPLICANT:  Bloksberg, Leonard
;  APPLICANT:  Lubbers, Mark
;  APPLICANT:  Dekker, James
;  APPLICANT:  Christensson, Anna
;  APPLICANT:  Holland, Ross
;  APPLICANT:  O'Toole, Paul
;  APPLICANT:  Reid, Julian
;  APPLICANT:  Coolbear, Timothy
;  TITLE OF INVENTION: Lactobacillus rhamnosus Polynucleotides, Polypeptides and Meth
;  TITLE OF INVENTION:  Using Them
;  FILE REFERENCE: 1043c2
;  CURRENT APPLICATION NUMBER: US/09/971,536
;  CURRENT FILING DATE:  2001-10-02
;  PRIOR APPLICATION NUMBER: U.S. No. 09/634,238
;  PRIOR FILING DATE: 2000-08-08
;  PRIOR APPLICATION NUMBER: U.S. No. 09/724,623
;  PRIOR FILING DATE: 2000-11-28
;  PRIOR APPLICATION NUMBER: PCT/NZ01/00160
;  PRIOR FILING DATE: 2001-08-08
;  NUMBER OF SEQ ID NOS: 83
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 70
;   LENGTH: 1879
;   TYPE: PRT
;   ORGANISM: Lactobacillus rhamnosus
US-09-971-536-70


  Query Match             10.7%;  Score 182;  DB 5;  Length 1879;
  Best Local Similarity   26.6%;  Pred. No. 0.0049;
  Matches  110;  Conservative   35;  Mismatches  154;  Indels  114;  Gaps   19;

Qy       4 DHD-----NHKGTVKNAV-DMAKAAEEAASAASAATG-NAAIGDVVKNSGAAAKGGEAAS 56
           |||     | | : | || | |  |::|  ||  |   :  |  | :|   |    |:|
Db    1252 DHDATLTANEKASQKQAVTDEATKAKKAIDAAKQADAVDQAKTDGIKAIDAQHHSGQALD 1311
```

```
Qy      57 VNGIAKGIKGIVDAAGKADAKEGKLDATGAEGTTNVNAGKLFVK--RAADDGGDADDAGK 114
                  |||||:  :||  |: |  ::      |  :||   | | :| |
Db    1312 DR--------------KADAKQ-VIDAEAAKVTAAIDQDNTLTKAQKAAQKQGVATEADK 1356

Qy     115 AAAAVAASAATGNAAIGDVVNGDVAKAKGGDAASVNGIAKGIKGIVDAAEKA-------- 166
           |  |: |:        ||    | ||  |  |             :||  ||
Db    1357 AKQAIDAA--------GDADAVDQAKTAGIQA------------IDAQHKAGKTIDSRH 1395

Qy     167 -DAKEG----------------KLNAAGAEG-----TTNADAGKLFVKNAGNV------- 197
            |||:                | ||  |       | ||  |  :  ||:
Db    1396 DDAKQAIDEEAAKVIKAIDQDPTLTAAQKEAQKQAVATEADKAKKAIDAAGDADAVDQAK 1455

Qy     198 ------------GGEAGDAGK--AAAAVAAVSGEQILKAIVH-AAKDGGEKQGKK--AAD 240
                       |:  || |  |  |: | :| ::  || | |     :|: :|   ||
Db    1456 TAGIKAIDEQHKSGQTVDARKEDAKKAIDAEAG-KVTDAIDHDATLTAAQKEAQKQAVAD 1514

Qy     241 RTNPIDAAIGGAGDNDA--AAAFATMKKDDQIAAAMVLRGMAKDGQ---FALKDAAAAHE 295
             :    ||  ||: ||   |  | :|  |:           | ||       ||  | :
Db    1515 EADKAKKAIDAAGNADAIDQAKSAGIKAIDE---------QHKSGQSIDTRKDDAKKAID 1565

Qy     296 GTVKNAVDIIKAAAEAASAASAATGSAAIGDVVNGNGATAKGGDAKSVNGIAK 348
           | |   | |         |  |    | | :      |    ||| :|:   |
Db    1566 GEVAKITDAIDHDPTLTDAEKATQKQAVIAEADKAKKAIDAAGDADAVDQAQK 1618


RESULT   7
US-09-500-746-17
; Sequence 17, Application US/09500746
; GENERAL INFORMATION:
;  APPLICANT: Winchester , Robert J.
;  APPLICANT:  Gulko, Percio
;  APPLICANT:  Seki, Tetsunori
;  TITLE OF INVENTION: USES OF INHIBITORS FOR THE ACTIVATION OF CXCR4 RECEPTOR BY SDF-1
;  TITLE OF INVENTION:  TREATING RHEVMATOID ARTHRITIS
;  FILE REFERENCE: 0575/57005-B
;  CURRENT APPLICATION NUMBER: US/09/500,746
;  CURRENT FILING DATE:  2000-02-09
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 17
;   LENGTH: 410
;   TYPE: PRT
;   ORGANISM: Human
US-09-500-746-17


  Query Match            10.1%;  Score 172.5;  DB 5;  Length 410;
  Best Local Similarity   27.5%;  Pred. No. 0.0036;
  Matches  102;  Conservative   11;  Mismatches  161;  Indels   97;  Gaps   13;

Qy       3 NDHDNHKGTVKNAVDMAKAAEEAASAASAATGNAAIGDVVKNSGAAAKG--GEAASVNGI 60
             |  :|: |   |  :    |  |  | :|  |     |    | ||   |  |  |   |
Db      58 NGGNNNAGANANTGNNTGGAGAANGACAANGGGGGNGTCGNNGGAGCNGNTGTGAGTGGG 117

Qy      61 AKGIKGIVDAAG--KADAKEGKLDATGAEGT------------------TNVNAGKLFV 99
           |  |  |  :| |   |:|  |   |  |  |                  || | |
Db     118 AAGAAGGCNACGTCAANAAGGACGAATATTTGCAANGNNGNNCAGGGCTGTNCNCG---- 173

Qy     100 KRAADDGGDADDAGKAAAAVAASAATGNAAIGDVVNGDVAKAKGGDAASVNGIAKGIKGI 159
                 || |      ||| ||:||  ||| |   ||    |  |  |:  |   |  |
Db     174 ------GGCAGTTTGTAAAAAAAAAAANAA-GAACNG--CGACAGACAAGTGTNNGTTGA 224

Qy     160 V---DAAEKADA-KEGKLNAAGAEGTTNADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVS 215
               |   |:|   |  | ||  ||      |      |||          ||  ||
Db     225 CCCGAAGCNANAGTGGATNCAGGAGTACCTGG------AGNNAACTATGAACAANTAAGC 278

Qy     216 GEQILKAIVHAAKDGGEKQGKKAADRTNPIDAAIG--------GAGDNDAAAAFATMKKD 267
           |         ||  ||       |  ||       |        |||   ||   |
Db     279 GCAACAGCCAAAGAGG-------ACTTNCCGCTAGACCCACTCGAGGAAAACTAA----- 326

Qy     268 DQIAAAMVLRGMAKDGQFALKDAAAAHEGTVKNAVDIIKAAAEAASAASAATGSAAIGDV 327
                               ||    ||          |   |: |:|    || |
Db     327 ----------------------AACCTTGT---------GAGAGATGAAAGGNCAAAGAC 355

Qy     328 VNGNGATAKGG 338
           | |  | ||
Db     356 GTGGGGGAGGG 366


RESULT   8
US-09-249-585A-3
; Sequence 3, Application US/09249585A
; GENERAL INFORMATION:
;  APPLICANT: Horlick, Robert
;  TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISOMES
;  FILE REFERENCE: 0867/0D905
;  CURRENT APPLICATION NUMBER: US/09/249,585A
;  CURRENT FILING DATE:  1999-02-11
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
;   LENGTH: 641
;   TYPE: PRT
;   ORGANISM: Epstein Barr Virus
US-09-249-585A-3


  Query Match            10.0%;  Score 171.5;  DB 5;  Length 641;
  Best Local Similarity   27.2%;  Pred. No. 0.0064;
  Matches   85;  Conservative    3;  Mismatches  151;  Indels   73;  Gaps    9;

Qy      27 SAASAATGNAAIGDVVKNSGAAAKGGEAASVNGIAKGIKGIVDAAGKADAKEGKLDATGA 86
           : | |  | |  |      || | ||   :      |  |   | | |    |     ||
Db      89 TGAGAGAGGAGAGGAGAGGGAGAGGGAGGAGGAGCAGAGGGAGAGGGA----GGAGGAGA 144

Qy      87 EGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGDVVNGDVAKAKGGDA 146
            |      |     |  || |  || | |   | |  |    |           || |
Db     145 GGGAGAGGG----AGGAGAGGGAGGAGGAGAGGGAGAGGGAGGAG----------AGGGA 190

Qy     147 ASVNGIAKGIKGIVDAAEKADAKEGKLNAAGAEGTTNADAGKLFVKNAGNVGGEAGDAGK 206
               |                |   | || |   | ||          || || ||
Db     191 GGAGGAGAG--------------GGAGAGGAGGAGGAGAG------GAGAGGGAGGAGG 229

Qy     207 AAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAADRTNPIDAAIGGAGDNDAAAAFATMKK 266
           | |  |  |           |  ||   |                |||   |  | |
Db     230 AGAGGAGAGG----------AGAGGAGAG-----------GAGGAGAGGAGGAGAG--- 264

Qy     267 DDQIAAAMVLRGMAKDGQFALKDAAAAHEGTVKNAVDIIKAAAEAASAASAATGSAAIGD 326
               |     |  |  |  |  |  |     |      |  |  | :   | |  |
Db     265 ----GAGGAGAGGGAGGAGAGGGAGGAGAGGAGA-----GAGGAGGAGAGGAGGAGAGG 315

Qy     327 VVNGNGATAKGG 338
           || | ||
Db     316 GAGAGGAGAGGG 327


RESULT   9
US-09-976-594-616
; Sequence 616, Application US/09976594
; GENERAL INFORMATION:
;  APPLICANT: Furness, Michael
;  APPLICANT:  Buchbinder, Jenny
;  TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROI
;  FILE REFERENCE: PA-0041 US
;  CURRENT APPLICATION NUMBER: US/09/976,594
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: 60/240,409
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 1143
;  SOFTWARE: PERL Program
; SEQ ID NO 616
;   LENGTH: 605
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
```

```
;   OTHER INFORMATION: Incyte ID No: 1692213CB1
;   NAME/KEY: unsure
;   LOCATION: 596
;   OTHER INFORMATION: unknown or other
US-09-976-594-616


  Query Match             10.0%;  Score 171;  DB 5;  Length 605;
  Best Local Similarity   30.2%;  Pred. No. 0.0064;
  Matches  103;  Conservative    7;  Mismatches  163;  Indels   68;  Gaps   12;

Qy      11 TVKNAVDMAKAAEEAASAASAATGNAAIGDVVKNSGAAAKGGEAASVNGIAKGIKGIVDA 70
           |   |   | |    |:||: ||| ||        |  | || ||:         |   |
Db     164 TGTGAAAAACATGGCAAAATCATGGAA--------GCCATGGAAAA---------GTTAA 206

Qy      71 AGKADAKEGKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAI 130
           ||          |||| ||    |      || ||    |    |     ||  ||
Db     207 AGTCTGG----GATGAGATTCAATG------AAGTGGCCGCACAGTATAGTGAAGATAAA 256

Qy     131 GDVVNGDVAKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGAEG------TTNA 184
           |   | |   ||  |   |   | |   | | |  |   | |  |      |      |
Db     257 GCCAGGCAAGGGGGTGACTTG--GGTTGGATGACCAGAGGGTCCATGGTGGGACCATTTC 314

Qy     185 DAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAA---DR 241
           |            | || || |       :|     |   |   ||   |  ||
Db     315 AA-----------GAAGCAGCATTTGCCTTGCCTGTA---AGTGGGATGGATAAGCCTG 359

Qy     242 TNPIDAAIGG-----AGDNDAAAAFATMKKDDQIAAAMVLRGMAKDGQFALKDAAAAHEG 296
           |   | |        |   || |           |||    | |          |  | |
Db     360 TGTTTACAGACCCACCGGTTAAGA---------CAAAATTTGGATATCATATTATTATGG 410

Qy     297 TVKNAVDIIKAAAEAASAASAATGSAAIGDVVNGNGATAKG 337
           |   |      |||  |:||: || : |      |  ||| |
Db     411 TCGAAGGAAGAAAATAAAATCATATGAAAGACTG-AATAAG 450


RESULT  10
US-09-952-267-7
; Sequence 7, Application US/09952267
; GENERAL INFORMATION:
;  APPLICANT: HANSEN, ERIC J.
;  APPLICANT:  AEBI, CHRISTOPH
;  APPLICANT:  COPE, LESLIE D.
;  APPLICANT:  MACIVER, ISOBEL
;  APPLICANT:  FISKE, MICHAEL J.
;  APPLICANT:  FREDENBURG, ROSS A.
;  TITLE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS
;  FILE REFERENCE: AMCY:024
;  CURRENT APPLICATION NUMBER: US/09/952,267
;  CURRENT FILING DATE:  2001-09-12
;  PRIOR APPLICATION NUMBER: 09/336,447
;  PRIOR FILING DATE: 1999-06-21
;  NUMBER OF SEQ ID NOS: 98
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
;   LENGTH: 624
;   TYPE: PRT
;   ORGANISM: Moraxella catarrhalis
US-09-952-267-7


  Query Match              9.8%;  Score 167.5;  DB 5;  Length 624;
  Best Local Similarity   25.4%;  Pred. No. 0.01;
  Matches   89;  Conservative   42;  Mismatches  145;  Indels   75;  Gaps   17;

Qy       9 KGTVKNAV---DMAKAAEEAASAASAATG--NAAIGDVVKNSGAAAKGGEAASVNGIAKG 63
           ||  | ||    :   |   |: ||| ||  ||| |:::   ||   ||      |
Db      23 KGHTKKAVLGSLLIVGALGMATTASAQTGSTNAANGNIISGVGAYVGGGVINQAKGNYPT 82

Qy      64 IKGIVD--AAGKADAKEGKLD--ATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAV 119
           : |  |  | |       |  |  | |   |                ||::: |    : |
Db      83 VGGGFDNRATGNYSVISGGFDNQAKGEHSTI--------------AGGESNQATGRNSTV 128

Qy     120 AASAATGNAAIGDVVNGDVAKAKGGDAASVNGIAKGIKGIVD-----AAEKADAKEGKLN 174
           |    : | |:|   |  ||       |   |  | |:    :     | | :   |  :
Db     129 A--GGSNNQAVG--TNSTVAGGSNNQAKGANSFAAGVGNQANTDNAVALGKNNTINGNNS 184

Qy     175 AA-GAEGTTNADAGKLFVKNAGNVGGEAGD-------AGKAAAAV--AAVSG------EQ 218
           || |:| | | :   :|:  :     ::|        :|| |  ||  | |:|
Db     185 AAIGSENTVNENQKNVFILGSNTTNAQSGSVLLGHETSGKEATAVSRARVNGLTLKNFSG 244

Qy     219 ILKA---IVHAAKDGGEKQ------GKKAADRTNPIDAAIGGAGDNDAAAAFATMKKDDQ 269
           : ||    |     | |:|      |: : | |:    |: |:       | ||   |:|
Db     245 VSKADNGTVSVGSQGKERQIVHVGAGQISDDSTD----AVNGS----QLYALATAVDDNQ 296

Qy     270 IAAAM----------VLRGMAKDGQFALKDAAAAHEGTVKNAVDIIKAAAE 310
           :          ::|: |:     :|  : |:    |  || |  |:
Db     297 YDIEINQDNIKDLQKEVKGLDKEVGVLSRDIGSLHDDVADNQADIAKNKAD 347


RESULT  11
US-09-952-267A-7
; Sequence 7, Application US/09952267A
; GENERAL INFORMATION:
;  APPLICANT: HANSEN, ERIC J.
;  APPLICANT:  AEBI, CHRISTOPH
;  APPLICANT:  COPE, LESLIE D.
;  APPLICANT:  MACIVER, ISOBEL
;  APPLICANT:  FISKE, MICHAEL J.
;  APPLICANT:  FREDENBURG, ROSS A.
;  TITLE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS
;  FILE REFERENCE: AMCY:024
;  CURRENT APPLICATION NUMBER: US/09/952,267A
;  CURRENT FILING DATE:  2001-09-12
;  PRIOR APPLICATION NUMBER: US/09/336,447A
;  PRIOR FILING DATE: 1999-06-21
;  NUMBER OF SEQ ID NOS: 98
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
;   LENGTH: 624
;   TYPE: PRT
;   ORGANISM: Moraxella catarrhalis
US-09-952-267A-7


  Query Match              9.8%;  Score 167.5;  DB 5;  Length 624;
  Best Local Similarity   25.4%;  Pred. No. 0.01;
  Matches   89;  Conservative   42;  Mismatches  145;  Indels   75;  Gaps   17;

Qy       9 KGTVKNAV---DMAKAAEEAASAASAATG--NAAIGDVVKNSGAAAKGGEAASVNGIAKG 63
           ||  | ||    :   |   |: ||| ||  ||| |:::   ||   ||      |
Db      23 KGHTKKAVLGSLLIVGALGMATTASAQTGSTNAANGNIISGVGAYVGGGVINQAKGNYPT 82

Qy      64 IKGIVD--AAGKADAKEGKLD--ATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAV 119
           : |  |  | |       |  |  | |   |                ||::: |    : |
Db      83 VGGGFDNRATGNYSVISGGFDNQAKGEHSTI--------------AGGESNQATGRNSTV 128

Qy     120 AASAATGNAAIGDVVNGDVAKAKGGDAASVNGIAKGIKGIVD-----AAEKADAKEGKLN 174
           |    : | |:|   |  ||       |   |  | |:    :     | | :   |  :
Db     129 A--GGSNNQAVG--TNSTVAGGSNNQAKGANSFAAGVGNQANTDNAVALGKNNTINGNNS 184

Qy     175 AA-GAEGTTNADAGKLFVKNAGNVGGEAGD-------AGKAAAAV--AAVSG------EQ 218
           || |:| | | :   :|:  :     ::|        :|| |  ||  | |:|
Db     185 AAIGSENTVNENQKNVFILGSNTTNAQSGSVLLGHETSGKEATAVSRARVNGLTLKNFSG 244

Qy     219 ILKA---IVHAAKDGGEKQ------GKKAADRTNPIDAAIGGAGDNDAAAAFATMKKDDQ 269
           : ||    |     | |:|      |: : | |:    |: |:       | ||   |:|
Db     245 VSKADNGTVSVGSQGKERQIVHVGAGQISDDSTD----AVNGS----QLYALATAVDDNQ 296

Qy     270 IAAAM----------VLRGMAKDGQFALKDAAAAHEGTVKNAVDIIKAAAE 310
           :          ::|: |:     :|  : |:    |  || |  |:
Db     297 YDIEINQDNIKDLQKEVKGLDKEVGVLSRDIGSLHDDVADNQADIAKNKAD 347


RESULT  12
PCT-US01-20545-4
```

```
; Sequence 4, Application PC/TUS0120545
; GENERAL INFORMATION:
;  APPLICANT: NEW YORK UNIVERSITY
;  TITLE OF INVENTION: MYCOBACTERIAL PROTEINS AS EARLY ANTIGENS FOR SERODIAGNOSIS  AND V
;  FILE REFERENCE: 32004-173355
;  CURRENT APPLICATION NUMBER: PCT/US01/20545
;  CURRENT FILING DATE:  2001-06-17
;  PRIOR APPLICATION NUMBER: US 60/212,048
;  PRIOR FILING DATE: 2000-06-16
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
;   LENGTH: 588
;   TYPE: PRT
;   ORGANISM: Mycobacterium tuberculosis
PCT-US01-20545-4
```

```
  Query Match             9.6%;  Score 163.5;  DB 1;  Length 588;
  Best Local Similarity   27.2%;  Pred. No. 0.016;
  Matches   77;  Conservative   15;  Mismatches  128;  Indels   63;  Gaps   11;

Qy      22 AEEAASAASAATGNAAIGDVVKNSGAAAKGGEAAS-----------------VNGIAKG 63
           |  : : | |    || ||   | ||   ||:|||                   || | |
Db     294 ANTSIANAGAGGNGAAGGDGGANGGAGGAGGQAASAGSSVGGDGGNGGAGGTGTNGHAGG 353

Qy      64 IKGIVDAAGK--------------ADAKEGKLDATGAEGTTNVNAG---KLFVKRAADDG 106
             |   | |:               |    | :  ||  |    | |    |  :    ||
Db     354 AGGAGGAGGRGGWLVGNGGNGGNGAAGGNGAIGGTGGAGGVPANQGGNSALGTQPVGGDG 413

Qy     107 GDADDAGKAA----AAVAASAATGNAA---IGDVVNGDVAKAKGGDAASVNGIAKGIKGI 159
           ||  : |               |   | |:   :|:  ||      |      ||   || |
Db     414 GDGGNGGTGGTGGRGGDGGSGGAGGASGWLMGNGGNGGNGGTGGSGGVGGNG---GIGG- 469

Qy     160 VDAAEKADAKEGK---LNAAGAEGTTNADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSG 216
             | |   :|        :| |  |    |||       |  ||: || |      |   |
Db     470 -DGAGGGNATSTSSIPFDAHGGNGGAGGDAGH------GGTGGDGGDGGH-----AGTGG 517

Qy     217 EQILKAIVHAAKDGGEKQGKKAADRTN--PIDAAIGGAGDNDA 257
             | |  ||    |   |   |  |:  |     || |  :|
Db     518 RGGLLAGQHANSGNGGGGGTGGAGGTHGTPGSGNAGGTGTGNA 560
```

```
RESULT  13
US-09-861-597-2
; Sequence 2, Application US/09861597
; GENERAL INFORMATION:
; APPLICANT: PHILLIPPE, Michel
; APPLICANT:  GARSON, Jean-Claude
;  APPLICANT:  ARRAUDEAU, Jean-Pierre
;  TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT
;  TITLE OF INVENTION:  LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN
;  TITLE OF INVENTION:  ANALOG
;  FILE REFERENCE: 6388-0365-0
;  CURRENT APPLICATION NUMBER: US/09/861,597
;  CURRENT FILING DATE:  2001-05-22
;  PRIOR APPLICATION NUMBER: 09/247,806
;  PRIOR FILING DATE: 1999-02-11
;  PRIOR APPLICATION NUMBER: FR 98/01614
;  PRIOR FILING DATE: 1998-02-11
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 529
;   TYPE: PRT
;   ORGANISM: Nephila clavipes
US-09-861-597-2
```

```
  Query Match             9.3%;  Score 158.5;  DB 5;  Length 529;
  Best Local Similarity   25.6%;  Pred. No. 0.028;
  Matches   96;  Conservative   21;  Mismatches  171;  Indels   87;  Gaps   12;

Qy      19 AKAAEEAASAASAATGNAAIGDVVKNSGAAAKGGEAASVNGIAKGIKGIVDAAGKADAKE 78
           | ||  |||| |   |    |     |    |    |   | :|     | | |
Db     112 AAAASAAASAESGGPGGYGPG----QQGPGGYGPGQQGPGGYGPGQQG-PSGPGSAAAAA 166

Qy      79 GKLDATGAEGTTNVNAGKLFVKRAADDG---GDADDAGKAAAAVAASAATGNAAIGDVVN 135
                | :|      |     :    |    |    :|  :|| ||:||:|    |
Db     167 AAASGPGQQGPGGYGPG-----QQGPGGYGPGQQGPSGPGSAAAAAAAASGPGQQGPGGY 221

Qy     136 GDVAKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAA-------GAEGTTNADAGK 188
           |   :  ||          :|: |   ||  | | |:           |  |  :| |
Db     222 GPGQQGPGGYGPG----QQGLSGPGSAAAAAAAGPGQQGPGGYGPGQQGPSGPGSAAAAA 277

Qy     189 LFVKNAGNVG-----------GEAG--DAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQG 235
                | |              |: |   || |||| ||  |:| |          ||   |
Db     278 AAAAGPGGYGPGQQGPGGYGPGQQGPSGAGSAAAAAAAGPGQQGL---------GGYGPG 328

Qy     236 KKAADRTNPIDAAIGGAGDNDAAAAFATMKKDDQIAAAMVLRGMAKDGQFALKDAAAAHE 295
           ::       |      || |   |:|| |       |         |  : |      |:||
Db     329 QQGPGGYGPGQQGPGGYGPGSASAAAAAAGPGQQGPGGY---GPGQQGPSGPGSASAA-- 383

Qy     296 GTVKNAVDIIKAAAEAA-------------------------SAASAATGSAAIGDVVNG 330
                     ||| ||                          :|||| :|| |    |
Db     384 -----------AAAAAAGPGGYGPGQQGPGGYAPGQQGPSGPGSASAAAAAAAAGPGGYG 432

Qy     331 NGATAKGGDAKSVNG 345
           |    || |    |
Db     433 PGQQGPGGYAPGQQG 447
```

```
RESULT  14
US-09-995-493-44
; Sequence 44, Application US/09995493
; GENERAL INFORMATION:
;  APPLICANT: Handfield, Martin
;  APPLICANT:  Hillman, Jeffrey
;  APPLICANT:  Progulske-Fox, Ann
;  TITLE OF INVENTION: Identification of Actinobacillus actinomycetemcomitans Antigen
;  TITLE OF INVENTION:  in the Diagnosis, Treatment, and Monitoring of Periodontal Di
;  FILE REFERENCE: MBHB01-662
;  CURRENT APPLICATION NUMBER: US/09/995,493
;  CURRENT FILING DATE:  2001-11-28
;  NUMBER OF SEQ ID NOS: 234
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 44
;   LENGTH: 1531
;   TYPE: PRT
;   ORGANISM: Actinobacillus actinomycetemcomitans
US-09-995-493-44
```

```
  Query Match             9.2%;  Score 157.5;  DB 5;  Length 1531;
  Best Local Similarity   20.7%;  Pred. No. 0.095;
  Matches  100;  Conservative   53;  Mismatches  157;  Indels  173;  Gaps   20;

Qy       2 NNDHDNHKGTVKNAVDMAKAAEEAASAASAATGNAAI---------------------- 38
           ::::|| ||        |||    |    | || ||:
Db     493 DSNYDN-KG--------AKAVGSIALGEKATTGTAAMNSIAIGLNSNVSGQNTVALGANI 543

Qy      39 -----GDVVKNSGAAAKGG------EAASVNGIA-KGIKGIVDAAG-------KADAKEG 79
                | |:  : :  :|        :|:|||    |  | |  :|       | : ::
Db     544 TATTNGSVILGNSSTTEGSHPVSNVSSATVNGYTYSGFTGTVKESGHFVSIGSKGNERQI 603

Qy      80 KLDATG---AEGTTNVNAGKLF-----------VKRAADDGGDADDAGKAAAAVAAS--- 122
           |  | |   |  |  ||  :||             :    ::||  : |    |   ::
Db     604 KNVAAGNVAANSTDAVNGSQLFAVASRVEQGWQITSGVENGGTQNGAASTATIKPSNQVK 663

Qy     123 ---------------------------------AATGNAAIGD--VVNGDVAKAKGGDA 146
                                             ||||  : :  | ||
Db     664 LLAGKNLAVKQNGTNFTFSTQENVTFTNVTTQDLTATGNTTVKNFSVQNGGTINMGNN-- 721

Qy     147 ASVNGIAKG-----------IKGIVDAAEKADAKEGKLNAAGAEGTTNADAGKLFVKNAG 195
             : |:|:|           :| ::  :   :  | |   || |    ||      ||||
Db     722 -RITGVAEGTQDDDAVNFKQLKSLLGGSASTEIVEKKAAQAGDENL--ADISVANGKNAG 778
```

```
Qy    196 NVGGE---------AGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAADRTNPID 246
          ::| :           | | |  |   :   : |  |:             | :: |:
Db    779 DMGAKYEVSVSKKAVQSAAKEAVKVTGSAPINVNKTDVNGVDTYAVTFNGTEAAKSIPLT 838

Qy    247 AAIGGAGD---------------------NDAAAAF----ATMKKDDQIAAAMVLRGM- 279
              |:||                     ||    :    :|:| :|  ||     |
Db    839 YKANGSGDKTVMLDKGLNFTNGMMTTASVANDGVVKYDVNLSTIKVEDGKAAVAGTPGTN 898

Qy    280 ---AKDGQFALKDAAAAHEGTVKNAVDIIKAAAEAASAASAATGSAAIGDVVNGNGATAK 336
               ||    ||  |    |||| |: :       :||   | | : |:||:    : |
Db    899 GANGTDG----KDGVA----TVKNVVEALN------NAAWTITASKSDGEVVSNASNSVK 944

Qy    337 GGD 339
           ||
Db    945 NGD 947


RESULT  15
US-09-646-673A-172
;  Sequence 172, Application US/09646673A
;  GENERAL INFORMATION:
;  APPLICANT: SPECHT, THOMAS
;  APPLICANT:  HINZMANN, BERND
;  APPLICANT:  SCHMITT, ARMIN
;  APPLICANT:  PILARSKY, CHRISTIAN
;  APPLICANT:  DAHL, EDGAR
;  APPLICANT:  ROSENTHAL, ANDRE
;  TITLE OF INVENTION: HUMAN NUCLEIC ACID SEQUENCES FROM TISSUE OF BREAST
;  TITLE OF INVENTION:  TUMORS
;  FILE REFERENCE: SCH-1763
;  CURRENT APPLICATION NUMBER: US/09/646,673A
;  CURRENT FILING DATE:  2000-09-20
;  PRIOR APPLICATION NUMBER: PCT/DE99/00908
;  PRIOR FILING DATE: 1999-03-19
;  NUMBER OF SEQ ID NOS: 202
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 172
;   LENGTH: 389
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-646-673A-172


  Query Match             9.1%;  Score 155;  DB 5;  Length 389;
  Best Local Similarity   27.9%;  Pred. No. 0.032;
  Matches   97;  Conservative   32;  Mismatches  145;  Indels   74;  Gaps   15;

Qy     13 KNAVDMAKAAE---EAASAASAATGNAAIGDVVKNSG----AAAKGG--EAASVNGIAKG 63
          | |     |||   ||| |:| :  |      || :|    |||: |  |    ||
Db     64 KTAAKGEAAAERPGEAAVASSPSKANGQENGHVKVNGDASPAAAESGAKEELQANG---- 119

Qy     64 IKGIVDAAGKADAKEGKLDATGAEGTTNVNAGKLFVKRAADDGG-----DADDAGKAAAA 118
                :|  || :|     :||   ::   |:     | : |       :|   |:||
Db    120 ------SAPAADKEEPAAAGSGAASPSSAEKGEPAAAAAPEAGASPVEKEAPAEGEAAEP 173

Qy    119 VAASAATGNAAIGDVVNGDVAKAKGGDAASVNGIAK------------GIKGIVDAAEKA 166
           :|:|| | ||           |   :  |   |              : |      |
Db    174 GSATAAEGEAASAASSTSSPKAEDGATPSPSNETPKKKKKRFSFKKSFKLSGFSFKKNKK 233

Qy    167 DAKE-GKLNAAGAEGTTNADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQILKAIVH 225
          :| | |:  |  |||       ||   : ||      | :|| |:   ||  ||:
Db    234 EAGEGGEAEAPAAEG------GK--DEAAGGAAAAAAEAGAASGEQAAAPGEEAAAGEEG 285

Qy    226 AAKDGGEKQGKKAADRTNPIDAAIG----GAGDNDAAA---AFATMKKDDQI---AAAMV 275
          ||  ||: |  |       | :||:       | |   ||   :    || ::    |||
Db    286 AA--GGDPQEAK------PQEAAVAPEKPPASDETKAAEEPSKVEEKKAEEAGASAAACE 337

Qy    276 LRGMAKDGQFALKDAAAAHEGTVKNAVDIIKAAAEAASAASAATGSAA 323
            |  |    ::|| | |           || |||:| ||    |
Db    338 APSAAGPGAPPEQEAAPAEE-----------PAAAAASSACAAPSQEA 374
```

Search completed: January 10, 2002, 14:08:51
Job time: 320 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 10, 2002, 14:02:26 ; Search time 24.93 Seconds
(without alignments)
1066.383 Million cell updates/sec

Title: US-09-445-803-2
Perfect score: 1708
Sequence: 1 KNNDHDNHKGTVKNAVDMAK.....GNGATAKGGDAKSVNGIAKG 349

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	266.5	15.6	353	2	I40300	outer membrane lip
2	250	14.6	2639	2	T31328	fibroin - Chinese
3	225.5	13.2	555	2	S70791	vsaB protein - Myc
4	224.5	13.1	362	2	I40304	outer membrane lip
5	219.5	12.9	246	2	A43579	Vmp7 protein homol
6	215.5	12.6	364	2	S11981	variable major pro
7	213.5	12.5	485	2	A60610	circumsporozoite p
8	208	12.2	1329	2	E70917	hypothetical glyci
9	205	12.0	1079	2	B70807	hypothetical glyci
10	204.5	12.0	398	1	OZZQAS	circumsporozoite p
11	204.5	12.0	1901	2	F70806	hypothetical glyci
12	202	11.8	1660	2	A70869	hypothetical glyci
13	201.5	11.8	429	2	A54504	circumsporozoite p
14	197	11.5	718	2	A36068	major ampullate fi
15	197	11.5	1381	2	E70806	hypothetical glyci
16	197	11.5	1489	2	D70807	hypothetical glyci
17	195	11.4	1538	2	H70846	hypothetical glyci
18	194.5	11.4	973	2	C85693	probable membrane
19	193	11.3	837	2	E70835	hypothetical glyci
20	193	11.3	1011	2	F70620	hypothetical glyci
21	192	11.2	369	2	S11980	variable major pro
22	192	11.2	694	2	F70868	hypothetical glyci
23	191.5	11.2	798	2	T21369	hypothetical prote
24	191.5	11.2	1156	2	T34852	probable secreted
25	190.5	11.2	1122	2	G64887	probable tail fibe
26	186	10.9	882	2	B70812	hypothetical glyci
27	186	10.9	1207	2	T23754	hypothetical prote
28	185.5	10.9	606	2	H70816	hypothetical glyci
29	185.5	10.9	1306	2	A70934	hypothetical glyci
30	184.5	10.8	615	2	H70589	hypothetical glyci
31	183.5	10.7	801	2	F70824	hypothetical glyci
32	182	10.7	778	2	F70963	hypothetical glyci
33	182	10.7	2055	2	T31110	extracellular matr
34	181.5	10.6	421	2	JV0057	tolA protein - Esc
35	181	10.6	532	2	F70580	hypothetical glyci
36	180.5	10.6	741	2	G70917	hypothetical glyci
37	179.5	10.5	767	2	E70895	hypothetical glyci
38	179	10.5	864	1	EART	elastin precursor
39	179	10.5	1822	2	S33441	EF protein - Strep
40	177	10.4	860	1	EAMS	elastin precursor
41	175.5	10.3	572	2	T08509	trbL protein - Ent
42	175	10.2	498	2	C70720	hypothetical glyci
43	174.5	10.2	853	2	A70896	hypothetical glyci
44	173.5	10.2	591	2	B70523	hypothetical glyci
45	173.5	10.2	595	2	A48658	flagellin - Escher

ALIGNMENTS

RESULT 1
I40300
outer membrane lipoprotein - Borrelia hermsii
C;Species: Borrelia hermsii
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 08-Oct-1999
C;Accession: I40300
R;Restrepo, B.I.; Kitten, T.; Carter, C.J.; Infante, D.; Barbour, A.G.
Mol. Microbiol. 6, 3299-3311, 1992
A;Title: Subtelomeric expression regions of Borrelia hermsii linear plasmids are high
A;Reference number: I40300; MUID:93133110
A;Accession: I40300
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-353 <RES>
A;Cross-references: GB:L04788; NID:g144023; PIDN:AAA22963.1; PID:g144024
C;Genetics:
A;Gene: vmp17

Query Match 15.6%; Score 266.5; DB 2; Length 353;
Best Local Similarity 33.0%; Pred. No. 1.8e-07;
Matches 114; Conservative 36; Mismatches 122; Indels 73; Gaps 15;

```
Qy     44 NSGAAAKGGEAASVNGIAKGIKGIVD----------AAG-KADAKEGKLDA--TGAEGT 89
          ||| |: :  : || || :|          | | ||| |: :   | | |
Db     20 NSGGVAEDPKTVYLTSIANLGKGFLDVFVTFGDMVTGAFGIKADTKKSDIGKYFTDIEST 79

Qy     90 -TNV----------NAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGDVV-NGD 137
           |:|          |     || | |:       :  |  |:  |||:  ||: | |||
Db     80 MTSVKKKLQDEVAKNGNYPKVKTAVDEFVAILGKIEKGAKEASKGATGDVIIGNTVKNGD 139

Qy    138 VAKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGAEGTTNADAGKLFVKNAGNV 197
           |  |:| ||| : |||| ||        |||| :|   :  :  | ||||
Db    140 ---AVPGEATSVNSLVKGIKEIVGVV----LKEGKADADATKDDSKKDIGKLFTATTDAN 192

Qy    198 GGEAGDAGKAAAAVAAVSGEQILKAIVHA-------AKDGGEKQGKKAADRTNPIDAAIG 250
            :  | |||:: ||:|  ||:||| :       : || ||        |:  : |:
Db    193 RADNAAAQAAAASIGAVTGADILQAIVQSKENPVANSTDGIEK-------ATDAAEIAVA 245

Qy    251 GAGDN-----DAAAAFATMKKDDQIAAAMVLRGMAKDGQFALKDAAAAHEGTVKNAVDII 305
           | ||     | |      |||  ||| : || |||:| |::|:   |       ||  |
Db    246 PAKDNKKEIKDGA------KKDAVIAAGIALRAMAKNGTFSIKNNEDA-------AVTTI 292

Qy    306 KAAAEAASAASAATGSAAIGDVV-------NGNGATAKGGDAKSV 343
          :|| :|     :|   || : |       |   || |  | |||
Db    293 NSAAASAVNKILSTLIIAIRNTVDSGLKTINEALATVKQED-KSV 336
```

RESULT 2
T31328
fibroin - Chinese oak silkmoth
C;Species: Antheraea pernyi (Chinese oak silkmoth)

```
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C;Accession: T31328
R;Sezutsu, H.; Tamura, T.; Yukuhiro, K.
submitted to the EMBL Data Library, August 1998
A;Description: Characterization of the full length fibroin gene of a wild silkworm, Anth
A;Reference number: Z20995
A;Accession: T31328
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2639 <SEZ>
A;Cross-references: EMBL:AF083334; NID:g3450882; PID:g3450883; PIDN:AAC32606.1
C;Genetics:
A;Introns: 14/3


  Query Match             14.6%;  Score 250;  DB 2;  Length 2639;
  Best Local Similarity   31.3%;  Pred. No. 8.1e-06;
  Matches  108;  Conservative   25;  Mismatches  170;  Indels   42;  Gaps    8;

Qy      21 AAEEAASAASAATGNAAIGDVVKNSGAAAKGGEAASVNGIAKGIKGIVDAAGKADAKEGK 80
           ||  ||:||:| :|    ||    ||::|    ||:    |:   |   ||| | |
Db     893 AAAAAAAAAAAGSGAGGRGDGGYGSGSSAAAAAAAAAASAARR-AGHDSAAGSAAAAAAA 951

Qy      81 LDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAA--------IGD 132
             | | |      |         |||   |:  |||| ||:|| |: |          ||
Db     952 AAAAAASGAGGSGGG-----YGWGDGGYGSDSAAAAAAAAAAAAAGSGAGGAGGGYGWGD 1006

Qy     133 VVNGD---VAKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGAEGTTNADAGKL 189
             |     | |    ||: :| |:|  |       : |      || | |:     |
Db    1007 GGYGSDSAAAAAAAAAAAAAAG-ARGSGGYGGYGSDSAAAAAAAAAAAAAGSGAGGVGGG 1065

Qy     190 FVKNAGNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAADRTNPIDAAI 249
           :    |  | ::  |  |||| || ||           |||   |  ||       ||
Db    1066 YGWGDGGYGSDSAAAAAAAAAAAAGSG-------AGGRGDGGYGSGSSAAAAAAAAAAAA 1118

Qy     250 GGAGDNDAAAAFATMKKDDQIAAAMVLRGMA-----KDGQFALKDAAAAHEGTVKNAVDI 304
             || : || : |        |||    |        || :    ||||
Db    1119 RRAGHDRAAGSAAAAAAAAAAAAAAAGSGAGGSGGGYGWGDGGYGSDSAAAA---------- 1167

Qy     305 IKAAAEAASAASAATGSAAIGDVVNGNGATAKGGDAKSVNGIAKG 349
             ||| ||:||| | ||   |   : : | |    | :  |    |
Db    1168 -AAAAAAAAAASGAGGSGGYGGYGSDSAAAAAAAAAAAAGSGAG 1211


RESULT   3
S70791
vsaB protein - Mycoplasma pulmonis (fragment)
C;Species: Mycoplasma pulmonis
C;Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 07-Dec-1999
C;Accession: S70791
R;Bhugra, B.; Voelker, L.L.; Zou, N.; Yu, H.; Dybvig, K.
Mol. Microbiol. 18, 703-714, 1995
A;Title: Mechanism of antigenic variation in Mycoplasma pulmonis: interwoven, site-speci
A;Reference number: S70790; MUID:96414471
A;Accession: S70791
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-555 <BHU>
A;Cross-references: EMBL:U23947; NID:g780764; PIDN:AAB41030.1; PID:g780766
C;Genetics:
A;Gene: vsaB
A;Genetic code: SGC3


  Query Match             13.2%;  Score 225.5;  DB 2;  Length 555;
  Best Local Similarity   27.2%;  Pred. No. 3.8e-05;
  Matches  101;  Conservative   40;  Mismatches  162;  Indels   69;  Gaps   15;

Qy       1 KNNDHDNHKGTV-KNAVDMAKAAEEAASAASAATGNAAIGDVVKNSGAAAKGGEA----- 54
           |||  | : | : ||    | | |   : |    |:|  ||: ||:   |  ||
Db      76 KNNAGDANAGEMGKNNAGDANAGEMGKNNA----GDANAGDMGKNNAGDANAGEMGKNNA 131

Qy      55 --ASVNGIAKGIKGIVDAA--GKADAKEGKLDATGAEGTTNVNAGKLFVKRAAD------ 104
             |:   : |   | :|   || :| :       |     : |||::    | |
Db     132 GDANAGDMGKNNAGDANAGEMGKNNAGDANAGEMGKNNAGDANAGEM 191

Qy     105 ---DGGDAD--DAGKAAAAVAASAATGNAAIGDVVNGDVAKAKGGDAASVNGIAKGIKGI 159
             : |||:  | ||  |  | : |    ||   |::  |   || ||:   : |   |
Db     192 GKNNAGDANAGDMGKNNAGDANAGEMGKNNAGDANAGEMGKNNAGD-ANAGEMGKNNAGD 250

Qy     160 VDAAEKADAKEGKLNAA--GAEGTTNADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSGE 217
             :| |      |  ||   |     :|:||::   |||:    ||: ||  |  |
Db     251 ANAGEMGKNNAGDANAGEMGKNNAGDANAGEMGKNNAGD--ANAGEMGKNNAGDA----- 303

Qy     218 QILKAIVHAAKDGGEKQGKKAAD-------RTNPIDAAIGGAGDNDAAAAFA-TMKKDDQ 269
                     : ||     | |       : |  ||  |  | |:|  | |  | |::
Db     304 -----------NAGEMGKNNAGDANAGEMGKNNAGDANAGEMGKNNAGDANAGEMGKNN- 351

Qy     270 IAAAMVLRGMAKDGQFALKDAAAAHEGTV--KNAVDIIKAAAEAASAASAATGSAAIGDV 327
                  | |  |:    :|  |: | :   || |     | |         | |  |::
Db     352 -------AGDANAGEMGKNNAGDANAGEMGKNNAGD-----ANAGEMGKNNAGDANAGEM 399

Qy     328 VNGNGATAKGGD 339
             |   |  |:
Db     400 GKNNAGDANAGE 411


RESULT   4
I40304
outer membrane lipoprotein - Borrelia hermsii
C;Species: Borrelia hermsii
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 08-Oct-1999
C;Accession: I40304
R;Restrepo, B.I.; Kitten, T.; Carter, C.J.; Infante, D.; Barbour, A.G.
Mol. Microbiol. 6, 3299-3311, 1992
A;Title: Subtelomeric expression regions of Borrelia hermsii linear plasmids are high
A;Reference number: I40300; MUID:93133110
A;Accession: I40304
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-362 <RES>
A;Cross-references: GB:L04787; NID:g144029; PIDN:AAA22965.1; PID:g144030
C;Genetics:
A;Gene: vmp25


  Query Match             13.1%;  Score 224.5;  DB 2;  Length 362;
  Best Local Similarity   28.4%;  Pred. No. 2.9e-05;
  Matches   99;  Conservative   37;  Mismatches  148;  Indels   65;  Gaps   14;

Qy      20 KAAEEAASAASAATGNAAIGDVVKNSGAAAKGGEAASVNGIAKGIKGIVDAAGKADAKEG 79
           :| |     |:|||  :: :|:   | :|:       |     :  : |  |    |:
Db      30 QAVEAGKDGAAAATGGRSLSEVLMEVGKSAE-------NAFYSFMALVPDTLGLRVTKDT 82

Qy      80 KLDATGAEGTTNVNAGKLFVKRAADD------GGDADDAGKAAAAVAASAATGNA----- 128
           | :  |  |  |   |||   :|:|:         : : |     |||  ||   |
Db      83 KKNEVG--GYFNSLGGKL--GKASDELEEVAKKSEVEGAKDGPIAVAIRAAVDTAKTTLS 138

Qy     129 ----------AIGDVVN-GDVAKAKGGDAAS---VNGIAKGIKGIVDAAEKADAKEGKL- 173
                      |||    |:    : | |||   : |  | :|||| | |    : |
Db     139 TLKEHLESLKGIGDDDKVGEATSNQNGVAASTDELKGAFKALKGIVDTAGKEGVAKPKAG 198

Qy     174 -NAAGAEGTTNADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGE 232
              |      | |  |:     | | |    ||  |||| |:|||||::| :||  |   |:
Db     199 DTAVKIGNADNKDGAKVLAA-AANAGRAVGD--KAAAIVSAVSGEEMLASIV--ASQEGD 253

Qy     233 KQGKKAADRTNPIDAAIGGAGDNDAAAAFATMKKDDQIAAAMVLRGMAKDGQF------- 285
               ||| |    |      |  :|         |   :|  : || : | |:
Db     254 ADAALAADATAQTSALKFARGGGNAGQLAKEAAKAAAVAGGIALRSLVKGGKLAANNNDD 313

Qy     286 ----------ALKDAAAAHEG----TVKNAVDIIKAAAEAASAASAATG 320
                      |:     | ||    |||| ::  |   : | |   |||
Db     314 DKVVQSAGVTAVNKLLVAVEGIIKKTVKNVLEKAKGEIDKARAPK-ATG 361


RESULT   5
```

```
A43579
Vmp7 protein homolog - Borrelia hermsii
C;Species: Borrelia hermsii
C;Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 23-Mar-1993
C;Accession: A43579
R;Barbour, A.G.; Carter, C.J.; Burman, N.; Freitag, C.S.; Garon, C.F.; Bergstroem, S.
Infect. Immun. 59, 390-397, 1991
A;Title: Tandem insertion sequence-like elements define the expression site for variable
A;Reference number: A43579; MUID:91099991
A;Accession: A43579
A;Molecule type: DNA
A;Residues: 1-246 <BAR>
A;Cross-references: GB:Z11876
C;Comment: This spirochete eludes the immune system to cause the disease relapsing fever
C;Keywords: membrane protein

  Query Match             12.9%;  Score 219.5;  DB 2;  Length 246;
  Best Local Similarity   31.9%;  Pred. No. 3.7e-05;
  Matches   86;  Conservative   27;  Mismatches  100;  Indels   57;  Gaps    9;

Qy      87 EGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGD----VVNGDVAKAK 142
Db       4 EGNPNASATETAVKTLIDNTLD---------KIIEGSKTVSDAIGDASDPIANVGANNAA 54

Qy     143 GGDAASVNGIAKGIKGIVD--------AAEKADAKEGKLNAAGAEGTTNADAGKLFVKNA 194
Db      55 GVAGTGIESLTKGIKAIVDVVLGKEGNAEAGTDKKAEDLSARTAAG--NGEAGKLFAANA 112

Qy     195 GNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAADRTNPIDAAIGGAGD 254
Db     113 GDDANAKKVAADAAKAVGAVTGADILQAMV--------KNGANAA--------------- 149

Qy     255 NDAAAAFATMKKDDQIAAAMVLRGMAKDGQFALKDAAAAHEGTVKNAVDIIKAAAEAASA 314
Db     150 ADAAKANA---KDGTIAGAIALTVMAKGGKFAGPTADSA------DYVTVAKGAAVSAIT 200

Qy     315 ASAATGSAAIGDVVNGNGATAKGGDAKSVN 344
Db     201 KALDTLTIAIRKTIDAGLKTVK--EAMKIN 228


RESULT   6
S11981
variable major protein 21 - Borrelia hermsii
C;Species: Borrelia hermsii
C;Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 08-Oct-1999
C;Accession: S11981
R;[illegible]man, N.; Bergstroem, S.; Restrepo, B.I.; Barbour, A.G.
M[illegible]Microbiol. 4, 1715-1726, 1990
A;[illegible]le: The variable antigens Vmp7 and Vmp21 of the relapsing fever bacterium Borrelia
A;Reference number: S11980; MUID:91171872
A;Accession: S11981
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-364 <BUR>
A;Cross-references: GB:M57256; EMBL:X53927; NID:g144025; PIDN:AAB59031.1; PID:g144026

  Query Match             12.6%;  Score 215.5;  DB 2;  Length 364;
  Best Local Similarity   25.9%;  Pred. No. 8.5e-05;
  Matches   92;  Conservative   42;  Mismatches  110;  Indels  111;  Gaps   14;

Qy      16 VDMAKAAEEA----------ASAASAATGNAAIGDVVKNSGAAAKGGEAA-------- 55
Db      54 MEVGKSAENAFYSFLELVSDTLGFTAKSTTKKEDVGGYFNSLG--GKLGEASNELEQVAK 111

Qy      56 -SVNGIAKG------IKGIVDAAGKA-DAKEGKLDATGAEGTTNVNAGKLFVKRAADDGG 107
Db     112 NSEAGIEKNDASKNPIRSAVNAAKKTLEALKGYLDSLGTV----------------G 152

Qy     108 DADDAGKAAAAVAASAATGNAAIGDVVNGDVAKAKGGDAASVNGIAKGIKGIVDAAEKAD 167
Db     153 DSNPVGWAS----------NNAQGAAV----------DEAELKKAYKALKGIMDTAEGAG 192

Qy     168 A---KEGKLNAAGAEGTTNADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQILKAIV 224
Db     193 VARPEVGNIAVKVGNGTDNKDGAKIL---ATDGAAAVGDAGKAAAILTTVSGKEMLASIV 249

Qy     225 HAAKDGGEKQGKKAADRTNPIDAAIGGAG------DNDAAAAFATMKKDDQIAAAMVLRG 278
Db     250 NSTEDKAVKITGNVTVETTPLEFAVGGNGAHLSQNANSKAAA---------VAGGIALRS 300

Qy     279 MAKDGQFAL--KDAAAAHEG------------------TVKNAVDIIKAAAEAA 312
Db     301 LVKGGKLAADNNDDDKASQGVGITAANKLLVAVEDIIKKTVKNVLEKAKGEIDKA 355


RESULT   7
A60610
circumsporozoite protein precursor - Plasmodium brasilianum
C;Species: Plasmodium brasilianum
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 20-Aug-1999
C;Accession: A60610; A28615
R;di Giovanni, L.; Cochrane, A.H.; Enea, V.
Exp. Parasitol. 70, 373-381, 1990
A;Title: On the evolutionary history of the circumsporozoite protein in plasmodia.
A;Reference number: A60610; MUID:90214818
A;Accession: A60610
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-485 <DIA>
A;Experimental source: strain Columbian
R;Lal, A.A.; de la Cruz, V.F.; Collins, W.E.; Campbell, G.H.; Procell, P.M.; McCutcha
J. Biol. Chem. 263, 5495-5498, 1988
A;Title: Circumsporozoite protein gene from Plasmodium brasilianum. Animal reservoirs
A;Reference number: A28615; MUID:88186854
A;Accession: A28615
A;Molecule type: DNA
A;Residues: 93-485 <LAL>
A;Cross-references: GB:J03203; NID:g160212; PIDN:AAA29553.1; PID:g160213
C;Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology
C;Keywords: sporozoite; surface antigen; tandem repeat
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-485/Product: circumsporozoite protein #status predicted <MAT>
F;114-369/Region: 4-residue repeats
F;410-463/Domain: thrombospondin type 1 repeat homology <THR1>

  Query Match             12.5%;  Score 213.5;  DB 2;  Length 485;
  Best Local Similarity   31.4%;  Pred. No. 0.00014;
  Matches  107;  Conservative   28;  Mismatches  163;  Indels   43;  Gaps   17;

Qy       1 KNNDHDNHKGTVKNAVDMAKAAEEAASAASAATGNAAIGDVVKNSGAAAKGGEAASVNGI 60
Db     112 EGNDAGNDAGNAAGNAAGNAAGNAAGNAAGNAAGNAA-GNAAGNAAGNAAGNDAGNAAGN 170

Qy      61 AKGIKGIVDAAGKADAKEGKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVA 120
Db     171 AAG-----NAAGNAAGNAAGNAAGNAAGNA---AGNAAGNAAGNDAGNA--AGNAAGNAA 220

Qy     121 ASAATGNAAIGDVVNGDVAKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAG--A 178
Db     221 GNAA-GNAA-GNAAGNDAGNAAGNAAGNAAGNAAG-----NAAGNA-AGNAAGNAAGNAA 272

Qy     179 EGTTNADAGKLFVKNAGNVGGE---------AGDAGKAAAAVAA--VSGEQILKAIVHAA 227
Db     273 GNAAGNDAGNAAGNAAGNAAGNAAGNAAGNAAGNAAGNAAGNAAGNAAGNAAGNAA 332

Qy     228 KD-GGEKQGKKAADRT-NPIDAAIGGAGDNDAAAAFATMK---KDDQIAAAMVLRGMAKD 282
Db     333 GNAAGNAAGNAAGNAAGNAAGNAAGNAAGNAAGNEKAKNKDNKVDA----NTNKKD 388

Qy     283 GQFALKDAA-AAHEGTVKNAVDIIK-AAAEAASAASAATGS 321
Db     389 NQEENNDSSNGPSEEHIKNYLESIRNSITEEWSPCSVTCGS 429
```

```
RESULT   8
E70917
hypothetical glycine-rich protein Rv1450c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: E70917
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987
A;Accession: E70917
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1329 <COL>
A;Cross-references: GB:Z95844; GB:AL123456; NID:g3250713; PIDN:CAB09271.1; PID:g2131046
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv1450c
C;Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homology


  Query Match             12.2%;  Score 208;  DB 2;  Length 1329;
  Best Local Similarity   26.4%;  Pred. No. 0.00067;
  Matches   88;  Conservative   17;  Mismatches  162;  Indels   66;  Gaps   10;

Qy      22 AEEAASAASAATGNAAIGDVVKNSGAAAKGGEAASVNGIAKGIKGIVDAAGKADAKEGKL 81
           |  || |:  : ||   |     :||    |  | || | |  |:|   |       |
Db     384 AHGAAGASPTSGGNGGAG----GNGAHFSSGGKAGGNGGAGGAGGLVGNGG-----AGGA 434

Qy      82 DATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGDVVNGDVAKA 141
             || |             |   ||| :  |  |         | |  ||   |  |
Db     435 GGNGAPG------------APPSGGDPNGGGGGAGGAGGKGGDGGAQAGDGGAGG-AGG 480

Qy     142 KGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGAEGTTNADAGKLFV-------KNA 194
           |||:  :    | |: |:   |:  |  :|    ||  |      ||            |
Db     481 KGGNGGNGATGATGLNGLGAGADGTDGGKGGNGGAGGGGGAGGQGGKALAATHQDGSMGA 540

Qy     195 GNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAADRTNPIDAAIGGAGD 254
           |  ||  |  |         |    |       ||     |      |      |||||
Db     541 GGAGGNGGAGGMGG------DGGNGAKGTFDNGGDGVGGNGGNGGSR------GIGGAGG 588

Qy     255 NDAAAAFATMKKDDQIAAAMVLRGMAKDGQFALKDAAAAHEGTVKNAVDIIKAAAEAASA 314
           | : |                  ||       :  : ||  |      | |  |
Db     589 IGGAGSTA----------------GADGARGATPTSGGNGGTGGNG-----ANATVAGG 626

Qy     315 ASAATGSAAIGDVVNGNGATAKGGDAKSVNGIA 347
           |  | |    | :|   ||  ||||  :  |:|
Db     627 AGGAGGKGGNGGLVGNGGAGGKGGDGMA--GVA 657


RESULT   9
B70807
hypothetical glycine-rich protein Rv3512 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: B70807
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987
A;Accession: B70807
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1079 <COL>
A;Cross-references: GB:AL022022; GB:AL123456; NID:g3261554; PIDN:CAA17749.1; PID:g292444
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv3512
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo


  Query Match             12.0%;  Score 205;  DB 2;  Length 1079;
  Best Local Similarity   28.4%;  Pred. No. 0.0008;
  Matches   99;  Conservative   14;  Mismatches  150;  Indels   86;  Gaps   18;

Qy      10 GTVKNAVDMAKAAEEAASAA--SAATGNAAIGDVVKNSGAAAKGGEAASVNGIAKGIKGI 67
           ||     |  |       ||  :   |  | |    : ||   ||:  : ||   |
Db     112 GTAGTGGDGGKGGNGGIGAAGTTGPVGTGASGGTGGSGGAGGTGGDGGAANGGTAG---- 167

Qy      68 VDAAGKADAKEGKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGN 127
              || |    ||    |  | |:  ||       |   ||   ||| |  | |   | |
Db     168 ---AGGAGGNGGK-GGDGGAGVTSSTAGN--SGGAGGSGGKGGDAG-AGGAGATPGANGI 220

Qy     128 AAIGDVVNGDVAKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGAEGTTNADAG 187
           | |   ||     ||| |:      |  ||  |    |   |   |||  | |    |
Db     221 AGNG----GD-----GGDGAA------GAVGISGATGAGDGGHGGTGAAGGNGGTGGAGG 265

Qy     188 KLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAADRTNPIDA 247
                 |  ||  |: |    |:    |:          |   |   |          :
Db     266 SGIDGVGGGTGGTGGNGGN--GAIGGAGGD---------AGGSGNSGG----------NG 304

Qy     248 AIGGAGDNDAAAAFATMKKDDQIAAAMVLRGMAK GQFALKDAAAAHEGTVKNAVDIIKA 307
           |||  | | | |  |       : |     |   ||     :  ||
Db     305 GIGGKGGN-AGAGGAAGSNGGTVGA----NGTGGDG----GNGGAA------------G 342

Qy     308 AAEAASAASAATGSAAIGDVVNGNGAT-------AKGGDAKSVNGIAKG 349
           || | |   | ||||       ||| |         |||   : |:  |
Db     343 AATAGSNGGAGTGSA------GGNGGTGGRGGSGGAGGD--GIGGVGGG 383


RESULT  10
OZZQAS
circumsporozoite protein precursor - Plasmodium cynomolgi (strain Ceylon)
N;Alternate names: major sporozoite surface antigen
C;Species: Plasmodium cynomolgi
C;Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 07-Nov-1997
C;Accession: C26255
R;Galinski, M.R.; Arnot, D.E.; Cochrane, A.H.; Barnwell, J.W.; Nussenzweig, R.S.; Ene
Cell 48, 311-319, 1987
A;Title: The circumsporozoite gene of the Plasmodium cynomolgi complex.
A;Reference number: A90889; MUID:87102878
A;Accession: C26255
A;Molecule type: DNA
A;Residues: 1-398 <GAL>
C;Comment: There are three distinct regions in the mature circumsporozoite protein, t
obic membrane-anchoring sequence.
C;Comment: There are 16 tandem copies of a 9-residue repeat and 3 copies of a 17-resi
C;Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology
C;Keywords: sporozoite; surface antigen; tandem repeat
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-398/Product: circumsporozoite protein #status predicted <MAT>
F;97-240/Region: 9-residue repeats
F;241-291/Region: 17-residue repeats
F;323-376/Domain: thrombospondin type 1 repeat homology <THR1>


  Query Match             12.0%;  Score 204.5;  DB 1;  Length 398;
  Best Local Similarity   34.8%;  Pred. No. 0.00035;
  Matches   86;  Conservative   14;  Mismatches  108;  Indels   39;  Gaps   10;

Qy      27 SAASAATGNAAIGDVVKNSGAAAKGGEAA---SVNGIAKGIKGIVDAAGKADAKEGKLDA 83
           :||  |  ||| |:   |: |   |  ||   : |  | |  |   |||:|    |   |
Db     101 AAAGEAGNNAAAGEAGNNAAAGEAGNNAAAGEAGNNAAAGEAGNNAAAGEA----GNNAA 156

Qy      84 TGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGDVVNGDVAKAKG 143
            || |  |  ||:      |    |:   ||:|    |  ||  ||| |:  |     | |
Db     157 GGAAG-NNAAAGEAGNNAAGGAAGNNAAAGEAGNNAAGGAAGNNAAAGEAGN----NAAG 211
```

```
Qy    144 GDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGAEGTTNADAGKLFVKNAGNVG----- 198
          |  |: |  | |  |   ||  |    |   |||| |  || ||:     ||  |
Db    212 G--AAGNNAAAGEAGNNAAAGAA----GNNAAAGAAG-NNAAAGEAGAGGAGRAGNNAAA 264

Qy    199 GEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAADRTNPIDAAIGGAGDNDAA 258
          ||||  |   |   | :||       |   ||    ||| |          | | |:
Db    265 GEAGAGGAGRAGNNAAAGE------AGAGGAGGNAGNKKAGD---------AGQGQNNGG 309

Qy    259 AAFATMK 265
          |    :|
Db    310 ANVPNVK 316


RESULT  11
F70806
hypothetical glycine-rich protein Rv3508 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: F70806
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
;[illegible]nor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
R[illegible]dream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987
A;Accession: F70806
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1901 <COL>
A;Cross-references: GB:AL022022; GB:AL123456; NID:g3261554; PIDN:CAA17745.1; PID:g292444
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv3508
C;Superfamily: collagen alpha 1(IV) chain


  Query Match             12.0%;  Score 204.5;  DB 2;  Length 1901;
  Best Local Similarity   27.6%;  Pred. No. 0.0014;
  Matches  106;  Conservative   25;  Mismatches  136;  Indels  117;  Gaps   19;

Qy     10 GTVKNA----------VDMAKAAEEAASAASAATGNAAIGDVVKNSGAAAKGGEAASVNG 59
          |:| ||          |  |  |  ||:| |:||| |          ||  :|| |   :|
Db    668 GSVGNAGIGGTGGTGGVGGAGGAGAAAAAGSSATGGAGFA-----GGAGGEGG-AGGNSG 721

Qy     60 IAKGIKGIVDAAGKADAKEGKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAV 119
          :    |  : :| |    ||    || |:   |          |   | |   | || |
Db    722 V-----GGTNGSGGAGGAGGKGGTGGAGGSGADN------PTGAGFAGGAGGTGGAAGAG 770

Qy    120 AASAATGNAAIGDVV--------------NGDVAKAKGGDAASVNGIAKGIKGIVDAAE- 164
          |  |||    | ||              ||     ||| ||  |:  |: | |  |  :
Db    771 GAGGATGTGGTGGVVGATGSAGIGGAGGRGGD-----GGDGAS--GLGLGLSGF-DGGQG 822

Qy    165 -----KADAKEGKLNAAGAEGTTNADAG-----------KLFVKNAGNVGGEAGDAGKAA 208
               |  | :| ||  |    | |                |   |  || ||: | |
Db    823 GQGGAGGSAGAGGINGAGGAGGNGGDGGDGATGAAGLGDNGGVGGDGGAGGAAGNGGNAG 882

Qy    209 AAVAAVSGEQILKAIVHAAKDGGEKQGKKAADRTNPIDAAIGGAGDNDAAAAFATMKKDD 268
          : | :|           |||       | :  |      ||||||:
Db    883 VGLTAKAG-----------DGG-----AAGNGGNGGAGGAGGAGDNNFNGG-------- 917

Qy    269 QIAAAMVLRGMA--KDGQFALKDAAA----AHEGTVKNAVDIIKAAAEAASAASAATGSA 322
                :| |  : ||  |  |:     |: |  |                  | |:
Db    918 --------QGGAGGQGGQGGLGGASTTSINANGGAGGNG---------GTGGKGGAGGAG 960

Qy    323 AIGDVVNGNGAT-AKGGDAKSVNG 345
          :|  | |:| |   |||| |  |
Db    961 TLG--VGGSGGTGGDGGDAGSGGG 982


RESULT  12
A70869
hypothetical glycine-rich protein Rv2490c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: A70869
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A;Reference number: A70500; MUID:98295987
A;Accession: A70869
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1660 <COL>
A;Cross-references: GB:AL021246; GB:AL123456; NID:g3261507; PIDN:CAA16067.1; PID:g279
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv2490c
C;Superfamily: collagen alpha 1(IV) chain


  Query Match             11.8%;  Score 202;  DB 2;  Length 1660;
  Best Local Similarity   26.5%;  Pred. No. 0.0017;
  Matches  114;  Conservative   18;  Mismatches  186;  Indels  112;  Gaps   17;

Qy      6 DNHKGTVKNAVDMAKAAEEAASAASAATG-NAAIGDVVKNSGAAAKGGEAASV------- 57
          |   |   :  |     :  |  | |  |   | |  | : |   |||  |
Db    633 DGDGGAGGDGGDPGAGGKGGAGGAGATEGVTGATGATVHSGGNGGKGGNGADATVAGANG 692

Qy     58 -NGIAKGIKGIV--------DAAGKADAKEGKLDATGAEGT--------TNVNAGKLFV- 99
            | | |  |:|          :| | |    :   ||:||         :  | |:  |
Db    693 GKGGAGGNGGLVGDGGAGGDGGSGAAGANGANVGEDGADGTLSGQPGEGSEANGGQGGVG 752

Qy    100 ----KRAADDGGDADDA---------GKAAAAVAASAATGNAAIGDVVNGD-------VA 139
              |  |||    |          | |  |  |  | |  |    |  |:||       |
Db    753 GGGAGGAGGDGGAGSSALGSGGNGGRGDAGQAGGAGGAGGAGGAGGSVSGDGGPGGKGGA 812

Qy    140 KAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGAEGTTNADAGKLFVKNAGNVGG 199
            ||  ||  |  ||  |  |:||          | |  |    | |   |   |  ||
Db    813 GGAGGAGASGGGGGKGASG-ADSAE----------AVGGAGGKGGDGG---VGGVGGDGG 858

Qy    200 EAGDAGKAAAA---------VAAVSGEQILKAIVHAAKDGGE-KQGKKAADRTNPIDAAI 249
            || |   ||          |  | |:  |        |||    |    |  :|    :
Db    859 PGGDGGAGGAAPAGQVGSHGVGGVGGDGGLGGAGGNGGDGGHGSDGGDGGDGGDPGAGGL 918

Qy    250 GGAG---DNDAAAAFATMKKDD--------------QIAAAMVLRGMAKDGQFA--LKD 289
            || |   |  ||                      | :|    |  ||  | : |
Db    919 GGLGGDSGNGTRAASGVDASDHGPGSGGNGGNGGNGAQASVAGGAGGNGGDGGNAGRVGD 978

Qy    290 AAAAHEGTVKNAVDIIKAAAEAASAASAATGSAAI--------------GDVVNGNGATA 335
            |   |          || |  | | | || |:                |      ||
Db    979 GGAGGNGG--------DGAAGANGANSGAPGSDALALGQPGGNGGQGDAGQAGGAGGAGG 1030

Qy    336 KGGDAKSVNG 345
            ||   ||:|
Db   1031 AGGAGGSVSG 1040


RESULT  13
A54504
circumsporozoite protein - Plasmodium malariae (strain Uganda 1-CDC)
C;Species: Plasmodium malariae
C;Date: 06-Oct-1994 #sequence_revision 04-Nov-1994 #text_change 20-Aug-1999
C;Accession: A54504
R;Lal, A.A.; de la Cruz, V.F.; Campbell, G.H.; Procell, P.M.; Collins, W.E.; McCutcha
Mol. Biochem. Parasitol. 30, 291-294, 1988
A;Title: Structure of the circumsporozoite gene of Plasmodium malariae.
A;Reference number: A54504; MUID:89040027
A;Accession: A54504
A;Molecule type: DNA
A;Residues: 1-429 <LAL>
A;Cross-references: GB:J03992; NID:g160220; PIDN:AAA29557.1; PID:g160221
C;Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology
```

```
C;Keywords: tandem repeat
F;354-407/Domain: thrombospondin type 1 repeat homology <THR1>


  Query Match             11.8%;  Score 201.5;  DB 2;  Length 429;
  Best Local Similarity   30.2%;  Pred. No. 0.00053;
  Matches   97;  Conservative   26;  Mismatches  135;  Indels   63;  Gaps   15;

Qy       3 NDHDNHKGTVKNAVDMAKAAEEAASAASAATGNAAIGDVVKNSGAAAKGGEAASVNGIAK 62
Db     114 NDAGNDAGNAAGNAAGNAAGNAAGNAAGNAAGNAA-GNAAGNAAGNAAGNDAGNAAGNAA 172

Qy      63 GIKGIVDAAGKADAKEGKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAAS 122
Db     173 G-----NAAGNAAGNAAGNDAGNAAGNAAGNAA----------GNAAGNAAGNAAGNAAG 217

Qy     123 AATGNAAIGDVVNGDVAKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGAEGTT 182
Db     218 NAAGNAA-GNAAGNDAGNAAGNAAGNAAGNAAG-----NAAGNA-AGNAAGNAAG----- 265

Qy     183 NADAGKLFVKNAGNVGGEAGDAGKAAA-AVAAVSGEQILKAIVHAAKDGGEKQGKKAADR 241
Db     266 NA-AGNAAGNAAGNAAGNA--AGNAAGNAAGNAAGNAAGNAAGNAA--GNAAGNEKAKNK 320

Qy     242 TNPIDAAIGGAGDNDAAAAFATMKKDDQIAAAMVLRGMAKDGQFALKDAAAAHEGTVKNA 301
Db     321 DNKVDA--------------NTNKKDNQ----------EENNDSSNGPSEEH---IKNY 352

Qy     302 VDIIK-AAAEAASAASAATGS 321
Db     353 LESIRNSITEEWSPCSVTCGS 373


RESULT  14
A36068
major ampullate fibroin protein - orb spider (Nephila clavipes) (fragment)
C;Species: Nephila clavipes
C;Date: 08-Mar-1991 #sequence_revision 13-Jan-1993 #text_change 09-Sep-1997
C;Accession: A36068
R;Xu, M.; Lewis, R.V.
Proc. Natl. Acad. Sci. U.S.A. 87, 7120-7124, 1990
A;Title: Structure of a protein superfiber: spider dragline silk.
A;Reference number: A36068; MUID:90384959
A;Accession: A36068
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-718 <XUA>
A;Cross-references: GB:M37137; NID:g159711; PID:g159712
A;Note: the authors translated the codon GGT for residue 292 as Gln, GTA for residue 374


  Query Match             11.5%;  Score 197;  DB 2;  Length 718;
  Best Local Similarity   26.1%;  Pred. No. 0.0014;
  Matches   98;  Conservative   32;  Mismatches  164;  Indels   82;  Gaps   13;

Qy      20 KAAEEAASAASAATGNAAIGDV-VKNSGAAAKGGE-----AASVNGIAKGIKGIVDAAGK 73
Db     234 QGAGQGAGASAAAAGGAGQGGYGGLGSQGAGRGGEGAGAAAAAAGGAGQGGYGGLGGQGA 293

Qy      74 ADAKEGKLDATGA--EGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASA-------- 123
Db     294 GQGGYGGLGSQGAGRGGLGGQGAGAAAAGGAGQGGLGGQGAGQGAGAAAAAAGGAGQGGY 353

Qy     124 ------ATGNAAIGDVVNGDVAKAKGGDAA-------SVNGIAKGIKGIVDAAEKADAKE 170
Db     354 GGLGSQGAGRGGLGGQGAGAVAAAAAGGAGQGGYGGLGSQGAGRGGQG----AGAAAAAA 409

Qy     171 GKLNAAGAEGTTNADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDG 230
Db     410 GGAGQRGYGGLGNQGAGR------GGLGGQG--AGAAAAAAAGGAGQGGYGGL---GNQG 458

Qy     231 GEKQGKKAADRTNPIDAAIGGAGDN----------------DAAAAFATMKKDDQIAAA 273
```

```
Db     459 AGRGGQGAA-------AAAGGAGQGGYGGLGSQGAGRGGQGAGAAAAAAVGAGQEGIRG- 510

Qy     274 MVLRGMAKDGQFALKDAAAAHEGTVKNAVDIIKAAAEAASAASAATGSAAIGDVVNGNGA 333
Db     511 ---QGAGQGGYGGLGSQGSGRGG----------LGGQGAGAAAAAAGGAGQGGLGGQGAG 557

Qy     334 TAKGGDAKSVNGIAKG 349
Db     558 QGAGAAAAAAGGVRQG 573


RESULT  15
E70806
hypothetical glycine-rich protein Rv3507 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: E70806
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A;Reference number: A70500; MUID:98295987
A;Accession: E70806
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1381 <COL>
A;Cross-references: GB:AL022022; GB:AL123456; NID:g3261554; PIDN:CAA17744.1; PID:g392
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv3507
C;Superfamily: collagen alpha 1(IV) chain


  Query Match             11.5%;  Score 197;  DB 2;  Length 1381;
  Best Local Similarity   29.1%;  Pred. No. 0.0026;
  Matches  107;  Conservative   14;  Mismatches  163;  Indels   84;  Gaps   16;

Qy      25 AASAASAATGNAAIGDVVKN----SGAAAKGGEAASVNG---------IAKGIKGIVDAA 71
Db     201 APGAMGGTGGNGGNGALLIGGGGLGGAGGMGGTGGGTGGTGGNGGNGALLIGAGGVGGAG 260

Qy      72 GKADAKEGKLDATGAEGT-TNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAI 130
Db     261 GIGGQGTGAGGAAGAGGTGGNGGAGGLFMN--GGDGGAGGQGGDGAAGDAAASAGGTGGK 318

Qy     131 GDVVNGDVAKAKGGDAASV---NGIAKGIKGIVDAAEKADAKEGKLNAAGAEGTTNADAG 187
Db     319 GG-QGGDGGTGGAGGAGPVLFGHGGAGGMGG--------QGGTGGMGGAGGDGTTVIAAG 369

Qy     188 KLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAADRTNPIDA 247
Db     370 ---TGGEGGTGGAAGAGGAAGARGALTSGG--LAGGVGAGGTGG---------------- 408

Qy     248 AIGGAGDNDAAAA------------FATMK-KDDQIAAAMVLRGMAKDGQFALK------ 288
Db     409 -TGGTGGNGADAAAVVGFGANGDPGFAGGKGGNGGIGGAAVTGGVAGDGGTGGKGGTGGA 467

Qy     289 -----DAAAAHEGTVKNAVDIIKAAAEAASAASAATGSAA--IGD----VVNGNGATAKG 337
Db     468 GGAGNDAGSTGNPGGKGGDGGIGGAGGAGGAAGTGNGGHAGNTGDGGDGGTGGNGGNGTG 527

Qy     338 GDAKSVNG 345
Db     528 G----VNG 531


Search completed: January 10, 2002, 14:04:28
Job time: 122 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 10, 2002, 14:04:31 ; Search time 15.37 Seconds
(without alignments)
832.533 Million cell updates/sec

Title: US-09-445-803-2
Perfect score: 1708
Sequence: 1 KNNDHDNHKGTVKNAVDMAK.....GNGATAKGGDAKSVNGIAKG 349

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	266.5	15.6	353	1	VM17_BORHE	P32777 borrelia he
2	224.5	13.1	362	1	VM25_BORHE	P32779 borrelia he
3	215.5	12.6	364	1	VM21_BORHE	P21875 borrelia he
4	213.5	12.5	393	1	CSP_PLABR	P14593 plasmodium
5	209.5	12.3	398	1	CSP_PLACC	P08673 plasmodium
6	204.5	12.0	1901	1	YZ08_MYCTU	O53553 mycobacteri
7	201.5	11.8	429	1	CSP_PLAMA	P13815 plasmodium
8	197	11.5	747	1	SPD1_NEPCL	P19837 nephila cla
9	192	11.2	369	1	VM07_BORHE	P21876 borrelia he
10	190.5	11.2	1120	1	STFR_ECOLI	P76072 escherichia
11	183.5	10.7	801	1	Y747_MYCTU	O53810 mycobacteri
12	182	10.7	778	1	YQ34_MYCTU	P71933 mycobacteri
13	181.5	10.6	421	1	TOLA_ECOLI	P19934 escherichia
14	179	10.5	864	1	ELS_RAT	Q99372 rattus norv
15	177	10.4	860	1	ELS_MOUSE	P54320 mus musculu
16	175	10.2	498	1	YI18_MYCTU	Q50615 mycobacteri
17	173.5	10.2	914	1	WA22_MYCTU	O06794 mycobacteri
18	171.5	10.0	488	1	CYL2_BOVIN	Q28092 bos taurus
19	171.5	10.0	641	1	EBN1_EBV	P03211 epstein-bar
20	170.5	10.0	774	1	STF_LAMBD	P03764 bacteriopha
21	169.5	9.9	419	1	CSP_PLACM	P08676 plasmodium
22	169.5	9.9	3591	1	FHAB_BORPE	P12255 bordetella
23	168	9.8	401	1	CSP_PLACG	P08674 plasmodium
24	166	9.7	957	1	Y278_MYCTU	P56877 mycobacteri
25	164	9.6	603	1	YD25_MYCTU	Q10637 mycobacteri
26	162.5	9.5	543	1	YP91_MYCTU	Q50630 mycobacteri
27	161.5	9.5	627	1	SPD2_NEPCL	P46804 nephila cla
28	161	9.4	542	1	PSGL_ONCMY	P12027 oncorhynchu
29	159	9.3	2038	1	FSH_DROME	P13709 drosophila
30	158.5	9.3	1654	1	OMPB_RICRI	Q53047 r outer mem
31	157.5	9.2	5263	1	FBOH_BOMMO	P05790 bombyx mori
32	156.5	9.2	515	1	YI40_MYCTU	Q50594 mycobacteri
33	155	9.1	331	1	MACS_HUMAN	P29966 homo sapien
34	154	9.0	1608	1	HLYA_SERMA	P15320 serratia ma
35	153.5	9.0	1150	1	APMU_PIG	P12021 sus scrofa
36	152.5	8.9	2249	1	OMPA_RICRI	P15921 rickettsia
37	150.5	8.8	280	1	MACS_CHICK	P16527 gallus gall
38	150	8.8	378	1	CSP_PLACL	P08675 plasmodium
39	150	8.8	747	1	ELS_BOVIN	P04985 bos taurus
40	149.5	8.8	463	1	YA68_MYCTU	O53416 mycobacteri
41	148.5	8.7	1655	1	OMPB_RICCN	Q9kka3 r outer mem
42	144.5	8.5	763	1	HTR2_HALN1	Q9hp81 halobacteri
43	144.5	8.5	764	1	HTR2_HALSA	P71410 halobacteri
44	144	8.4	1706	1	CYAA_BORPE	P15318 bordetella
45	143	8.4	3726	1	TRX_DROME	P20659 drosophila

ALIGNMENTS

RESULT 1
VM17_BORHE
ID VM17_BORHE STANDARD; PRT; 353 AA.
AC P32777;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE VARIABLE MAJOR OUTER MEMBRANE LIPOPROTEIN 17 PRECURSOR.
GN VMP17.
OS Borrelia hermsii.
OG Plasmid.
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=140;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SSP. HS1 SEROTYPE 17;
RX MEDLINE=93133110; PubMed=1484486;
RA Restrepo B.I., Kitten T., Carter C.J., Infante D., Barbour A.G.;
RT "Subtelomeric expression regions of Borrelia hermsii linear plasmids are highly polymorphic.";
RL Mol. Microbiol. 6:3299-3311(1992).
CC -!- FUNCTION: SERVES TO AVOID THE HOST IMMUNE RESPONSE BY CHANGING FROM ONE SURFACE EXPOSED VMP TO ANOTHER.
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID ANCHOR.
CC ---

CC ---
DR EMBL; L04788; AAA22963.1; -.
DR InterPro; IPR000680; Borrelia_lipo.
DR InterPro; IPR001800; Lipoprotein_6.
DR Pfam; PF00921; Lipoprotein_2; 1.
DR Pfam; PF01441; Lipoprotein_6; 1.
DR ProDom; PD001149; Lipoprotein_6; 1.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Outer membrane; Lipoprotein; Signal; Plasmid.
FT SIGNAL 1 18 PROBABLE.
FT CHAIN 19 353 VARIABLE MAJOR OUTER MEMBRANE
FT LIPOPROTEIN 17.
FT LIPID 19 19 N-ACYL DIGLYCERIDE (PROBABLE).
SQ SEQUENCE 353 AA; 36460 MW; 88FAB9B8C619AE91 CRC64;

Query Match 15.6%; Score 266.5; DB 1; Length 353;
Best Local Similarity 33.0%; Pred. No. 6.7e-07;
Matches 114; Conservative 36; Mismatches 122; Indels 73; Gaps 15;

```
Qy     44 NSGAAAKGGEAASVNGIAKGIKGIVD----------AAG-KADAKEGKLDA--TGAEGT 89
          |||  |:  :   :  ||   || :|           | | ||| |:  :    |  | |
Db     20 NSGGVAEDPKTVYLTSIANLGKGFLDVFVTFGDMVTGAFGIKADTKKSDIGKYFTDIEST 79
```

```
Qy      90 -TNV----------NAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGDVV-NGD 137
           |:|          |     || | |:        :  |  |:  |||:  ||: | |||
Db      80 MTSVKKKLQDEVAKNGNYPKVKTAVDEFVAILGKIEKGAKEASKGATGDVIIGNTVKNGD 139

Qy     138 VAKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGAEGTTNADAGKLFVKNAGNV 197
             |  |:| ||| : |||| ||        |||| :|   :  :  | ||||
Db     140 ---AVPGEATSVNSLVKGIKEIVGVV----LKEGKADADATKDDSKKDIGKLFTATTDAN 192

Qy     198 GGEAGDAGKAAAAVAAVSGEQILKAIVHA-------AKDGGEKQGKKAADRTNPIDAAIG 250
             :   |  |||:: ||:|  ||:||| :       : || ||        |:  : |:
Db     193 RADNAAAQAAAASIGAVTGADILQAIVQSKENPVANSTDGIEK-------ATDAAEIAVA 245

Qy     251 GAGDN-----DAAAAFATMKKDDQIAAAMVLRGMAKDGQFALKDAAAAHEGTVKNAVDII 305
           | ||     | |      |||  ||| : || |||:| |::|:   |       ||  |
Db     246 PAKDNKKEIKDGA------KKDAVIAAGIALRAMAKNGTFSIKNNEDA-------AVTTI 292

Qy     306 KAAAEAASAASAATGSAAIGDVV-------NGNGATAKGGDAKSV 343
           :|| :|     :|   || : |       |   || |  | |||
Db     293 NSAAASAVNKILSTLIIAIRNTVDSGLKTINEALATVKQED-KSV 336
```

RESULT 2
VM25_BORHE
ID VM25_BORHE STANDARD; PRT; 362 AA.
AC P32779;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE VARIABLE MAJOR OUTER MEMBRANE LIPOPROTEIN 25 PRECURSOR.
GN VMP25.
OS Borrelia hermsii.
OG Plasmid.
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=140;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SSP. HS1 SEROTYPE 25;
RX MEDLINE=93133110; PubMed=1484486;
RA Restrepo B.I., Kitten T., Carter C.J., Infante D., Barbour A.G.;
RT "Subtelomeric expression regions of Borrelia hermsii linear plasmids
RT are highly polymorphic.";
RL Mol. Microbiol. 6:3299-3311(1992).
CC -!- FUNCTION: SERVES TO AVOID THE HOST IMMUNE RESPONSE BY CHANGING
CC FROM ONE SURFACE EXPOSED VMP TO ANOTHER.
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID
CC ANCHOR.
CC -!- SIMILARITY: STRONG, TO VMP7 AND VMP21.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; L04787; AAA22965.1; -.
DR InterPro; IPR000680; Borrelia_lipo.
DR Pfam; PF00921; Lipoprotein_2; 1.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Outer membrane; Lipoprotein; Signal; Plasmid.
FT SIGNAL 1 26 PROBABLE.
FT CHAIN 27 362 VARIABLE MAJOR OUTER MEMBRANE
FT LIPOPROTEIN 25.
FT LIPID 27 27 N-ACYL DIGLYCERIDE (PROBABLE).
SQ SEQUENCE 362 AA; 36468 MW; 2A3EF5248D5637FD CRC64;

Query Match 13.1%; Score 224.5; DB 1; Length 362;
Best Local Similarity 28.4%; Pred. No. 7.3e-05;
Matches 99; Conservative 37; Mismatches 148; Indels 65; Gaps 14;

```
Qy      20 KAAEEAASAASAATGNAAIGDVVKNSGAAAKGGEAASVNGIAKGIKGIVDAAGKADAKEG 79
           :| |     |:||||  :: :|:   | :|:       |     :  : | |    |:
Db      30 QAVEAGKDGAAAATGGRSLSEVLMEVGKSAE-------NAFYSFMALVPDTLGLRVTKDT 82

Qy      80 KLDATGAEGTTNVNAGKLFVKRAADD------GGDADDAGKAAAAVAASAATGNA----- 128
           | :  |  |  |   |||   :|:|:        : : |     |||  ||   |
Db      83 KKNEVG--GYFNSLGGKL--GKASDELEEVAKKSEVEGAKDGPIAVAIRAAVDTAKTTLS 138

Qy     129 ----------AIGDVVN-GDVAKAKGGDAAS---VNGIAKGIKGIVDAAEKADAKEGKL- 173
                     |||    |:   : | |||   : |  | :|||| | |    : |
Db     139 TLKEHLESLKGIGDDDKVGEATSNQNGVAASTDELKGAFKALKGIVDTAGKEGVAKPKAG 198

Qy     174 -NAAGAEGTTNADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGE 232
            |       | |  |:    | | |   ||  |||| |:|||||::| :||  |   |:
Db     199 DTAVKIGNADNKDGAKVLAA-AANAGRAVGD--KAAAIVSAVSGEEMLASIV--ASQEGD 253

Qy     233 KQGKKAADRTNPIDAAIGGAGDNDAAAAFATMKKDDQIAAAMVLRGMAKDGQF------- 285
           |||  |    |     |  :|       |   :|  : || : | |:
Db     254 ADAALAADATAQTSALKFARGGGNAGQLAKEAAKAAAVAGGIALRSLVKGGKLAANNNDD 313

Qy     286 ----------ALKDAAAAHEG----TVKNAVDIIKAAAEAASAASAATG 320
                    |:     | ||    |||| ::  |   : | |   |||
Db     314 DKVVQSAGVTAVNKLLVAVEGIIKKTVKNVLEKAKGEIDKARAPK-ATG 361
```

RESULT 3
VM21_BORHE
ID VM21_BORHE STANDARD; PRT; 364 AA.
AC P21875;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE VARIABLE MAJOR OUTER MEMBRANE LIPOPROTEIN 21 PRECURSOR.
GN VMP21.
OS Borrelia hermsii.
OG Plasmid.
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=140;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SSP. HS1 SEROTYPE 21;
RX MEDLINE=91171872; PubMed=1706456;
RA Burman N., Bergstroem S., Restrepo B.I., Barbour A.G.;
RT "The variable antigens Vmp7 and Vmp21 of the relapsing fever
RT bacterium Borrelia hermsii are structurally analogous to the VSG
RT proteins of the African trypanosome.";
RL Mol. Microbiol. 4:1715-1726(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SSP. HS1 SEROTYPE 21;
RX MEDLINE=93133110; PubMed=1484486;
RA Restrepo B.I., Kitten T., Carter C.J., Infante D., Barbour A.G.;
RT "Subtelomeric expression regions of Borrelia hermsii linear plasmids
RT are highly polymorphic.";
RL Mol. Microbiol. 6:3299-3311(1992).
RN [3]
RP SEQUENCE OF 55-75; 186-208 AND 245-259.
RX MEDLINE=85236116; PubMed=2409197;
RA Barstad P.A., Coligan J.E., Raum M.G., Barbour A.G.;
RT "Variable major proteins of Borrelia hermsii. Epitope mapping and
RT partial sequence analysis of CNBr peptides.";
RL J. Exp. Med. 161:1302-1314(1985).
CC -!- FUNCTION: SERVES TO AVOID THE HOST IMMUNE RESPONSE BY CHANGING
CC FROM ONE SURFACE EXPOSED VMP TO ANOTHER.
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID
CC ANCHOR.
CC -!- SIMILARITY: STRONG, TO VMP7 AND VMP25.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M57256; AAB59031.1; -.
DR   PIR; S11981; S11981.
DR   InterPro; IPR000680; Borrelia_lipo.
DR   Pfam; PF00921; Lipoprotein_2; 1.
DR   PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW   Outer membrane; Lipoprotein; Signal; Plasmid.
FT   SIGNAL        1     26       PROBABLE.
FT   CHAIN        27    364       VARIABLE MAJOR OUTER MEMBRANE
FT                                LIPOPROTEIN 21.
FT   LIPID        27     27       N-ACYL DIGLYCERIDE (PROBABLE).
SQ   SEQUENCE   364 AA;  37084 MW;  16598B639EE63776 CRC64;


  Query Match             12.6%;  Score 215.5;  DB 1;  Length 364;
  Best Local Similarity   25.9%;  Pred. No. 0.0002;
  Matches   92;  Conservative   42;  Mismatches  110;  Indels  111;  Gaps   14;

Qy      16 VDMAKAAEEA------------ASAASAATGNAAIGDVVKNSGAAAKGGEAA-------- 55
           ::: |:|| |              | : |    :|    : |   | |||:
Db      54 MEVGKSAENAFYSFLELVSDTLGFTAKSTTKKEDVGGYFNSLG--GKLGEASNELEQVAK 111

Qy      56 -SVNGIAKG------IKGIVDAAGKA-DAKEGKLDATGAEGTTNVNAGKLFVKRAADDGG 107
            |  || |        |:  |:|| |  :| :| ||: |                    |
Db     112 NSEAGIEKNDASKNPIRSAVNAAKKTLEALKGYLDSLGTV------------------G 152

Qy     108 DADDAGKAAAAVAASAATGNAAIGDVVNGDVAKAKGGDAASVNGIAKGIKGIVDAAEKAD 167
           |::  | |:          | | |  |          | | :    | :|||:| || |
Db     153 DSNPVGWAS----------NNAQGAAV----------DEAELKKAYKALKGIMDTAEGAG 192

Qy     168 A---KEGKLNAAGAEGTTNADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQILKAIV 224
               : | :      || | |  |:    | :     |||||||| :  |||:::| :||
Db     193 VARPEVGNIAVKVGNGTDNKDGAKIL---ATDGAAAVGDAGKAAAILTTVSGKEMLASIV 249

Qy     225 HAAKDGGEKQGKKAADRTNPIDAAIGGAG------DNDAAAAFATMKKDDQIAAAMVLRG 278
           :: :|   |        | |:: |:|| |       |  |||         :|  : ||
Db     250 NSTEDKAVKITGNVTVETTPLEFAVGGNGAHLSQNANSKAAA---------VAGGIALRS 300

Qy     279 MAKDGQFAL--KDAAAAHEG------------------TVKNAVDIIKAAAEAA 312
           : | |: |    |   | :|                  ||| :: |   : |
Db     301 LVKGGKLAADNNDDDKASQGVGITAANKLLVAVEDIIKKTVKNVLEKAKGEIDKA 355


RESULT   4
CSP_PLABR
ID   CSP_PLABR      STANDARD;      PRT;   393 AA.
AC   P14593;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   CIRCUMSPOROZOITE PROTEIN (CS) (FRAGMENT).
OS   Plasmodium brasilianum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5824;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88186854; PubMed=3128542;
RA   Lal A.A., la Cruz V.F., Collins W.E., Campbell G.H., Procell P.M.,
RA   McCutchan T.F.;
RT   "Circumsporozoite protein gene from Plasmodium brasilianum. Animal
RT   reservoirs for human malaria parasites?";
RL   J. Biol. Chem. 263:5495-5498(1988).
CC   -!- FUNCTION: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT
CC       SURFACE ANTIGEN ON THE SPOROZOITE (THE INFECTIVE STAGE OF THE
CC       MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE
CC       VERTEBRATE HOST).
CC   -!- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR
CC       ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES
CC       WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.
CC   -!- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; J03203; AAA29553.1; -.
DR   PIR; A28615; A28615.
DR   HSSP; P04002; 1WFA.
DR   InterPro; IPR000884; TSP1.
DR   Pfam; PF00090; tsp_1; 1.
DR   SMART; SM00209; TSP1; 1.
DR   PROSITE; PS50092; TSP1; 1.
KW   Malaria; Repeat; Sporozoite.
FT   NON_TER       1      1
FT   DOMAIN       16    278       TANDEM REPEATS OF G-N-A-A.
SQ   SEQUENCE   393 AA;  35372 MW;  B32944419BC600AA CRC64;


  Query Match             12.5%;  Score 213.5;  DB 1;  Length 393;
  Best Local Similarity   31.4%;  Pred. No. 0.00027;
  Matches  107;  Conservative   28;  Mismatches  163;  Indels   43;  Gaps   17;

Qy       1 KNNDHDNHKGTVKNAVDMAKAAEEAASAASAATGNAAIGDVVKNSGAAAKGGEAASVNGI 60
           : ||  |  |           |   | :||  | |||| |:   |:   | | :| :  |
Db      20 EGNDAGNDAGNAAGNAAGNAAGNAAGNAAGNAAGNAA-GNAAGNAAGNAAGNDAGNAAGN 78

Qy      61 AKGIKGIVDAAGKADAKEGKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVA 120
           | |     :||| |         |  | |     ||       | :| |:|  || ||   |
Db      79 AAG-----NAAGNAAGNAAGNAAGNAAGNA---AGNAAGNAAGNDAGNA--AGNAAGNAA 128

Qy     121 ASAATGNAAIGDVVNGDVAKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAG--A 178
           :|| |||| |:    |   | | |  | :  | | |     :||  | |     ||||  |
Db     129 GNAA-GNAA-GNAAGNDAGNAAGNAAGNAAGNAAG-----NAAGNA-AGNAAGNAAGNAA 180

Qy     179 EGTTNADAGKLFVKNAGNVGGE---------AGDAGKAAAAVAA--VSGEQILKAIVHAA 227
               |||      |||  |           ||:|   ||  ||   :|     |  :||
Db     181 GNAAGNDAGNAAGNAAGNAAGNAAGNAAGNAAGNAAGNAAGNAAGNAAGNAAGNAA 240

Qy     228 KD-GGEKQGKKAADRT-NPIDAAIGGAGDNDAAAAFATMK---KDDQIAAAMVLRGMAKD 282
           :  |   | | :   |    | | |  | | |  |   ||::: |        ||
Db     241 GNAAGNAAGNAAGNAAGNAAGNAAGNAAGNAAGNEKAKNKDNKVDA----NTNKKD 296

Qy     283 GQFALKDAA-AAHEGTVKNAVDIIK-AAAEAASAASAATGS 321
           |    |::   |  :|| :: |: :  |  |  |   ||
Db     297 NQEENNDSSNGPSEEHIKNYLESIRNSITEEWSPCSVTCGS 337


RESULT   5
CSP_PLACC
ID   CSP_PLACC      STANDARD;      PRT;   398 AA.
AC   P08673;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   CIRCUMSPOROZOITE PROTEIN PRECURSOR (CS).
OS   Plasmodium cynomolgi (strain Ceylon).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5829;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87102878; PubMed=3802196;
RA   Galinski M.R., Arnot D.E., Cochrane A.H., Barnwell J.W.,
RA   Nussenzweig R.S., Enea V.;
RT   "The circumsporozoite gene of the Plasmodium cynomolgi complex.";
RL   Cell 48:311-319(1987).
CC   -!- MISCELLANEOUS: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT
CC       SURFACE ANTIGEN ON THE SPOROZOITE (THE INFECTIVE STAGE OF THE
CC       MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE
CC       VERTEBRATE HOST).
CC   -!- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR
```

```
CC       ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES
CC       WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.
CC   -!- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M15103; AAA29533.1; -.
DR   PIR; C26255; OZZQAS.
DR   InterPro; IPR003067; Crcmsprzoite.
DR   InterPro; IPR000884; TSP1.
DR   Pfam; PF00090; tsp_1; 1.
DR   PRINTS; PR01303; CRCMSPRZOITE.
DR   SMART; SM00209; TSP1; 1.
KW   Malaria; Sporozoite; Repeat; Signal.
FT   SIGNAL        1     19       PROBABLE.
FT   CHAIN        20    398       CIRCUMSPOROZOITE PROTEIN.
FT   DOMAIN       97    249       17 X 9 AA TANDEM REPEATS OF A-G-N-N-A-A-
FT                                A-G-E.
SQ   SEQUENCE   398 AA;  37718 MW;  6DFA2E8A62ED05BF CRC64;


  Query Match             12.3%;  Score 209.5;  DB 1;  Length 398;
  Best Local Similarity   27.2%;  Pred. No. 0.00042;
  Matches   97;  Conservative   21;  Mismatches  129;  Indels  109;  Gaps   11;

Qy       7 NHKGTVKNAVDMAKAAEEAASAASAATGNAAIGDVVKNSGAAAKGGEAASVNGIAKGIKG 66
Db      33 NLNGVSFNNVDASSLG--AAQVRQSASRGRGLGENPKNEEGA---------------- 72

Qy      67 IVDAAGKADAK--------EGKL------DATGAEGTTNVNAGKLFVKRAADDGGDADDA 112
Db      73 --DKPKKKDEKQVEPKKPRENKLKQPAGNNAAAGEAGNNAAAGEAGNNAAAGEAGNNAAA 130

Qy     113 GKAAAAVAASAATGNAAIGDVVNGDVAKAKGGDAA---SVNGIAKGIKGIVDAAEKADAK 169
Db     131 GEAGNNAAAGEAGNNAAAGEAGNNAAGGAAGNNAAGGEAGNNAAGGAAG-----NNAAAG 185

Qy     170 EGKLNAAGAEGTTNADAGKLFVKNAGNVGGE---AGDAGKAAAAVAAVSGEQILKAIVHA 226
Db     186 EAGNNAAGGEAGNNAAAGEAGNNAAGGAAGNNAAAGEAGNNAAAGAAGN---------- 234

Qy     227 AKDGGEKQGKKAADRTNPIDAAIGGAGDNDAAAAFATMKKDDQIAAAMVLRGMAKDGQFA 286
Db     235 ------------------NAAAGAAGNNAAAG-------------------------- 248

Qy     287 LKDAAAAHEGTVKNAVDIIKAAAEAASAA--SAATGSAAIGDVVNGNGATAKGGDA 340
Db     249 --EAGAGGAGRAGNNAAAGEAGAGGAGRAGNNAAAGEAGAGG-AGGNAGNKKAGDA 301


RESULT   6
YZ08_MYCTU
ID   YZ08_MYCTU     STANDARD;      PRT;  1901 AA.
AC   O53553;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   HYPOTHETICAL PE-PGRS FAMILY PROTEIN RV3508 PRECURSOR.
GN   RV3508 OR MTV023.15.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
CC   -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC       SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL022022; CAA17745.1; -.
DR   HSSP; P19972; 1KVD.
DR   TubercuList; Rv3508; -.
DR   InterPro; IPR000084; PE.
DR   Pfam; PF00934; PE; 1.
DR   ProDom; PD001223; PE; 1.
KW   Hypothetical protein; Repeat; Signal; Complete proteome.
FT   SIGNAL        1     30       POTENTIAL.
FT   CHAIN        31   1901       HYPOTHETICAL PE-PGRS FAMILY PROTEIN
FT                                RV3508.
SQ   SEQUENCE   1901 AA;  147627 MW;  C7B1923D5D0146CD CRC64;


  Query Match             12.0%;  Score 204.5;  DB 1;  Length 1901;
  Best Local Similarity   27.6%;  Pred. No. 0.0029;
  Matches  106;  Conservative   25;  Mismatches  136;  Indels  117;  Gaps   19;

Qy      10 GTVKNA----------VDMAKAAEEAASAASAATGNAAIGDVVKNSGAAAKGGEAASVNG 59
Db     668 GSVGNAGIGGTGGTGGVGGAGGAGAAAAAGSSATGGAGFA-----GGAGGEGG-AGGNSG 721

Qy      60 IAKGIKGIVDAAGKADAKEGKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAV 119
Db     722 V-----GGTNGSGGAGGAGGKGGTGGAGGSGADN------PTGAGFAGGAGGTGGAAGAG 770

Qy     120 AASAATGNAAIGDVV--------------NGDVAKAKGGDAASVNGIAKGIKGIVDAAE- 164
Db     771 GAGGATGTGGTGGVVGATGSAGIGGAGGRGGD-----GGDGAS--GLGLGLSGF-DGGQG 822

Qy     165 -----KADAKEGKLNAAGAEGTTNADAG----------KLFVKNAGNVGGEAGDAGKAA 208
Db     823 GQGGAGGSAGAGGINGAGGAGGNGGDGGDGATGAAGLGDNGGVGGDGGAGGAAGNGGNAG 882

Qy     209 AAVAAVSGEQILKAIVHAAKDGGEKQGKKAADRTNPIDAAIGGAGDNDAAAAFATMKKDD 268
Db     883 VGLTAKAG-----------DGG-----AAGNGGNGGAGGAGGAGDNNFNGG-------- 917

Qy     269 QIAAAMVLRGMA--KDGQFALKDAAA----AHEGTVKNAVDIIKAAAEAASAASAATGSA 322
Db     918 --------QGGAGGQGGQGGLGGASTTSINANGGAGGNG---------GTGGKGGAGGAG 960

Qy     323 AIGDVVNGNGAT-AKGGDAKSVNG 345
Db     961 TLG--VGGSGGTGGDGGDAGSGGG 982


RESULT   7
CSP_PLAMA
ID   CSP_PLAMA      STANDARD;      PRT;   429 AA.
AC   P13815;
DT   01-JAN-1990 (Rel. 13, Created)
```

```
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   CIRCUMSPOROZOITE PROTEIN PRECURSOR (CS).
OS   Plasmodium malariae.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5858;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89040027; PubMed=3054537;
RA   Lal A.A., la Cruz V.F., Campbell G.H., Procell P.M., Collins W.E.,
RA   McCutchan T.F.;
RT   "Structure of the circumsporozoite gene of Plasmodium malariae.";
RL   Mol. Biochem. Parasitol. 30:291-294(1988).
CC   -!- FUNCTION: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT
CC       SURFACE ANTIGEN ON THE SPOROZOITE (THE INFECTIVE STAGE OF THE
CC       MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE
CC       VERTEBRATE HOST).
CC   -!- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR
CC       ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES
CC       WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.
CC   -!- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J03992; AAA29557.1; -.
DR   PIR; A54504; A54504.
DR   HSSP; P04002; 1WFA.
DR   InterPro; IPR003067; Crcmsprzoite.
DR   InterPro; IPR000884; TSP1.
DR   Pfam; PF00090; tsp_1; 1.
DR   PRINTS; PR01303; CRCMSPRZOITE.
DR   SMART; SM00209; TSP1; 1.
DR   PROSITE; PS50092; TSP1; 1.
KW   Malaria; Sporozoite; Repeat; Signal.
FT   SIGNAL        1     15       PROBABLE.
FT   CHAIN        16    429       CIRCUMSPOROZOITE PROTEIN.
FT   DOMAIN      110    314       4 AA TANDEM REPEATS.
SQ   SEQUENCE   429 AA;  41596 MW;  3629D641D1C0BB7E CRC64;


  Query Match             11.8%;  Score 201.5;  DB 1;  Length 429;
  Best Local Similarity   30.2%;  Pred. No. 0.0011;
  Matches   97;  Conservative   26;  Mismatches  135;  Indels   63;  Gaps   15;

Qy       3 NDHDNHKGTVKNAVDMAKAAEEAASAASAATGNAAIGDVVKNSGAAAKGGEAASVNGIAK 62
           ||  |  |           |   | :||  | |||| |:   |:   | | :| :  | |
Db     114 NDAGNDAGNAAGNAAGNAAGNAAGNAAGNAAGNAA-GNAAGNAAGNAAGNDAGNAAGNAA 172

Qy      63 GIKGIVDAAGKADAKEGKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAAS 122
           |     :||| |       ||  | |    ||           |  | :|   ||  ||
Db     173 G-----NAAGNAAGNAAGNDAGNAAGNAAGNAA----------GNAAGNAAGNAAGNAAG 217

Qy     123 AATGNAAIGDVVNGDVAKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGAEGTT 182
            | |||| |:    |   | |  | :  | | |     :||  | |     ||||
Db     218 NAAGNAA-GNAAGNDAGNAAGNAAGNAAGNAAG-----NAAGNA-AGNAAGNAAG----- 265

Qy     183 NADAGKLFVKNAGNVGGEAGDAGKAAA-AVAAVSGEQILKAIVHAAKDGGEKQGKKAADR 241
           || ||      ||| | |  || ||  |    :|     |  :||  |     :|| ::
Db     266 NA-AGNAAGNAAGNAAGNA--AGNAAGNAAGNAAGNAAGNAAGNAA--GNAAGNEKAKNK 320

Qy     242 TNPIDAAIGGAGDNDAAAAFATMKKDDQIAAAMVLRGMAKDGQFALKDAAAAHEGTVKNA 301
            | :||               | |||:|          ::   :    :  |   :||
Db     321 DNKVDA-------------NTNKKDNQ-----------EENNDSSNGPSEEH---IKNY 352

Qy     302 VDIIK-AAAEAASAASAATGS 321
           :: |: :  |  |  |   ||
Db     353 LESIRNSITEEWSPCSVTCGS 373
```

```
RESULT   8
SPD1_NEPCL
ID   SPD1_NEPCL     STANDARD;      PRT;   747 AA.
AC   P19837;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   SPIDROIN 1 (DRAGLINE SILK FIBROIN 1) (FRAGMENT).
OS   Nephila clavipes (Orb spider).
OC   Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC   Araneomorphae; Entelegynae; Araneoidea; Tetragnathidae; Nephila.
OX   NCBI_TaxID=6915;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=90384959; PubMed=2402494;
RA   Xu M., Lewis R.V.;
RT   "Structure of a protein superfiber: spider dragline silk.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:7120-7124(1990).
RN   [2]
RP   SEQUENCE OF 653-747 FROM N.A.
RX   MEDLINE=94165058; PubMed=8120021;
RA   Beckwitt R., Arcidiacono S.;
RT   "Sequence conservation in the C-terminal region of spider silk
RT   proteins (Spidroin) from Nephila clavipes (Tetragnathidae) and
RT   Araneus bicentenarius (Araneidae).";
RL   J. Biol. Chem. 269:6661-6663(1994).
CC   -!- FUNCTION: THIS SPIDER MAJOR AMPULLATE SILK POSSESSES UNIQUE
CC       CHARACTERISTICS OF STRENGTH AND ELASTICITY. IT MAY CONSIST OF
CC       PSEUDOCRISTALLINE REGIONS OF ANTIPARALLEL BETA-SHEET INTERSPERSED
CC       WITH ELASTIC AMORPHOUS SEGMENTS.
CC   -!- SUBUNIT: MAJOR SUBUNIT, WITH SPIDROIN 2, OF THE DRAGLINE SILK.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M37137; AAA29380.1; -.
DR   EMBL; U03848; AAB60212.1; -.
DR   PIR; A36068; A36068.
KW   Silk; Repeat.
FT   NON_TER       1      1
FT   DOMAIN        1    655       25 X APPROXIMATE TANDEM REPEATS.
FT   REPEAT        1     25       1.
FT   REPEAT       26     38       2.
FT   REPEAT       39     66       3.
FT   REPEAT       67     96       4.
FT   REPEAT       97    130       5.
FT   REPEAT      131    158       6.
FT   REPEAT      159    191       7.
FT   REPEAT      192    204       8.
FT   REPEAT      205    235       9.
FT   REPEAT      236    262       10.
FT   REPEAT      263    292       11.
FT   REPEAT      293    305       12.
FT   REPEAT      306    333       13.
FT   REPEAT      334    360       14.
FT   REPEAT      361    394       15.
FT   REPEAT      395    424       16.
FT   REPEAT      425    458       17.
FT   REPEAT      459    485       18.
FT   REPEAT      486    512       19.
FT   REPEAT      513    525       20.
FT   REPEAT      526    555       21.
FT   REPEAT      556    582       22.
FT   REPEAT      583    612       23.
FT   REPEAT      613    642       24.
```

```
FT   REPEAT      643    655       25.
FT   CONFLICT    662    662       V -> L (IN REF. 1).
FT   CONFLICT    672    672       S -> T (IN REF. 1).
FT   CONFLICT    695    747       NPGLSGCDVLIQALLEVVSALIQILGSSSIGQVNYGSAGQA
FT                                TQIVGQSVYQAL -> ILVFLDVMSSFKLFSRLFLLLSRS
FT                                (IN REF. 1).
SQ   SEQUENCE   747 AA;  60528 MW;  850E44B0D649E012 CRC64;


  Query Match             11.5%;  Score 197;  DB 1;  Length 747;
  Best Local Similarity   26.1%;  Pred. No. 0.0029;
  Matches   98;  Conservative   32;  Mismatches  164;  Indels   82;  Gaps   13;

Qy      20 KAAEEAASAASAATGNAAIGDV-VKNSGAAAKGGE-----AASVNGIAKGIKGIVDAAGK 73
           : | : | |::|| | |  |       |  | :|||     ||:  |  :|  | :   |
Db     234 QGAGQGAGASAAAAGGAGQGGYGGLGSQGAGRGGEGAGAAAAAAGGAGQGGYGGLGGQGA 293

Qy      74 ADAKEGKLDATGA--EGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASA-------- 123
              | | : ||   |     ||       |   |      ||: | | ||:|
Db     294 GQGGYGGLGSQGAGRGGLGGQGAGAAAAGGAGQGGLGGQGAGQGAGAAAAAAGGAGQGGY 353

Qy     124 ------ATGNAAIGDVVNGDVAKAKGGDAA-------SVNGIAKGIKGIVDAAEKADAKE 170
                 |   :|    | || |  | |             |  :| :|    |  | |
Db     354 GGLGSQGAGRGGLGGQGAGAVAAAAAGGAGQGGYGGLGSQGAGRGGQG----AGAAAAAA 409

Qy     171 GKLNAAGAEGTTNADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDG 230
           |     |  |  |  ||:      | :||:   || |||| |  :|:     :      |
Db     410 GGAGQRGYGGLGNQGAGR------GGLGGQG--AGAAAAAAAGGAGQGGYGGL---GNQG 458

Qy     231 GEKQGKKAADRTNPIDAAIGGAGDN----------------DAAAAFATMKKDDQIAAA 273
           : |: ||       || ||||                    |||| |     : |
Db     459 AGRGGQGAA-------AAAGGAGQGGYGGLGSQGAGRGGQGAGAAAAAAVGAGQEGIRG- 510

Qy     274 MVLRGMAKDGQFALKDAAAAHEGTVKNAVDIIKAAAEAASAASAATGSAAIGDVVNGNGA 333
             :|  : |   |    :   |              : | ||:|| | |  | :
Db     511 ---QGAGQGGYGGLGSQGSGRGG----------LGGQGAGAAAAAAGGAGQGGLGGQGAG 557

Qy     334 TAKGGDAKSVNGIAKG 349
             |  | :  |: :|
Db     558 QGAGAAAAAAGGVRQG 573


RESULT   9
VM07_BORHE
ID   VM07_BORHE     STANDARD;      PRT;   369 AA.
AC   P21876;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   VARIABLE MAJOR OUTER MEMBRANE LIPOPROTEIN 7 PRECURSOR.
GN   VMP7.
OS   Borrelia hermsii.
OG   Plasmid.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=140;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SSP. HS1 SEROTYPE 7;
RX   MEDLINE=91171872; PubMed=1706456;
RA   Burman N., Bergstroem S., Restrepo B.I., Barbour A.G.;
RT   "The variable antigens Vmp7 and Vmp21 of the relapsing fever
RT   bacterium Borrelia hermsii are structurally analogous to the VSG
RT   proteins of the African trypanosome.";
RL   Mol. Microbiol. 4:1715-1726(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SSP. HS1 SEROTYPE 7;
RX   MEDLINE=93133110; PubMed=1484486;
RA   Restrepo B.I., Kitten T., Carter C.J., Infante D., Barbour A.G.;
RT   "Subtelomeric expression regions of Borrelia hermsii linear plasmids
RT   are highly polymorphic.";
RL   Mol. Microbiol. 6:3299-3311(1992).
RN   [3]
RP   SEQUENCE OF 52-72 AND 244-258.
RX   MEDLINE=85236116; PubMed=2409197;
RA   Barstad P.A., Coligan J.E., Raum M.G., Barbour A.G.;
RT   "Variable major proteins of Borrelia hermsii. Epitope mapping and
RT   partial sequence analysis of CNBr peptides.";
RL   J. Exp. Med. 161:1302-1314(1985).
CC   -!- FUNCTION: SERVES TO AVOID THE HOST IMMUNE RESPONSE BY CHANGING
CC       FROM ONE SURFACE EXPOSED VMP TO ANOTHER.
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID
CC       ANCHOR.
CC   -!- SIMILARITY: STRONG, TO VMP21 AND VMP25.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X53926; CAA37873.1; -.
DR   EMBL; Z11876; CAA77935.1; -.
DR   PIR; S11980; S11980.
DR   InterPro; IPR000680; Borrelia_lipo.
DR   Pfam; PF00921; Lipoprotein_2; 1.
DR   PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW   Outer membrane; Lipoprotein; Signal; Plasmid.
FT   SIGNAL        1     26       PROBABLE.
FT   CHAIN        27    369       VARIABLE MAJOR OUTER MEMBRANE
FT                                LIPOPROTEIN 7.
FT   LIPID         27     27       N-ACYL DIGLYCERIDE (PROBABLE).
SQ   SEQUENCE   369 AA;  37102 MW;  F318727E39399DA8 CRC64;


  Query Match             11.2%;  Score 192;  DB 1;  Length 369;
  Best Local Similarity   26.5%;  Pred. No. 0.0027;
  Matches   94;  Conservative   51;  Mismatches  136;  Indels   74;  Gaps   16;

Qy      10 GTVKNA-VDMAKAAEEAASA----ASAATGNAAIGDVVKN------SGAAAKGGEAAS-V 57
           | : |: ::: ::|| |  |     |   |  |  |  |       :   || |||:: :
Db      44 GNLGNSLMELGRSAENAFYAFIELVSDVLGFTAKSDTTKQEVGGYFNSLGAKLGEASNDL 103

Qy      58 NGIAKGIKGIVDAAGKADAKEGKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAA 117
             :|   :  ||   |:|: :  :     |    :   | :|: :    ||::  | |
Db     104 EQVAVKAETGVD---KSDSSKNPIREAVNEAKEVLGTLKGYVE-SLGTIGDSNPVGYANN 159

Qy     118 AVAASAATGNAAIGDVVNGDVAKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAG 177
           |    | :|  |  |    :: ||            | :: || ||  |  |  |: |
Db     160 A----AGSGTTAADD----ELRKA----------FKALQEIVKAATDAGVKALKIGATT 200

Qy     178 AE---GTTNADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQ 234
             :  |  | :  |:   : ||    | |  |||| :::|||::| :||       |
Db     201 LQANGGADNKEGAKILATSGGNPA--AADVAKAAAILSSVSGEEMLSSIV--------KS 250

Qy     235 GKKAADRTNPIDAAIGGAGD-NDAAAAFATMKKDDQIAAAMVLRGMAKDGQFALKDAAAA 293
           |:          || :  | | | :| :||    |  :| |   :  |  |  ||:
Db     251 GEN--------DAQLAAAADGNTSAISFAKGGSDAHLAGANTPKAAAVAGGIALR----- 297

Qy     294 HEGTVKNAVDIIKAAAEAASAASAATGSAAIGDVVNGNGATAKGGDAKSVNGIAK 348
                   ::|    || ||  |||    |  | | |  |     ::|  : |
Db     298 ---------SLVKTGKLAAGAADNATGG---GKEVQGVGVAAANKLLRAVEDVIK 340


RESULT  10
STFR_ECOLI
ID   STFR_ECOLI     STANDARD;      PRT;  1120 AA.
AC   P76072; P77560;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   SIDE TAIL FIBER PROTEIN HOMOLOG FROM LAMBDOID PROPHAGE RAC.
GN   STFR OR B1372.
OS   Escherichia coli.
```

```
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97251357; PubMed=9097039;
RA   Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA   Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA   Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA   Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,
RA   Sampei G., Seki Y., Sivasundaram S., Tagami H., Takeda J.,
RA   Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT   "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 28.0-40.1 min region on the linkage map.";
RL   DNA Res. 3:363-377(1996).
CC   -!- SIMILARITY: BELONGS TO THE TAIL FIBER FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000234; AAC74454.1; ALT_INIT.
DR   EMBL; D90774; BAA14966.1; -.
DR   EMBL; D90775; BAA14975.1; -.
DR   HSSP; P04002; 1WFA.
DR   EcoGene; EG13370; stfR.
DR   InterPro; IPR000122; Chemotaxis_transducer.
KW   Hypothetical protein; Fiber protein; Repeat; Complete proteome.
SQ   SEQUENCE   1120 AA;  113779 MW;  542E59D71EE795B4 CRC64;


  Query Match             11.2%;  Score 190.5;  DB 1;  Length 1120;
  Best Local Similarity   24.9%;  Pred. No. 0.0085;
  Matches   92;  Conservative   47;  Mismatches  168;  Indels   63;  Gaps   12;

Qy        6 DNHKGTVKNAVDMAKAAEEAASAASAATGNAAIGDVVKNSGAAAKGGEAASVNGIAKGIK 65
            :: |     :   || :|  |||:  :   :|     | | ||    :||:    ||  :
Db      191 ESSKSAAATSAGAAKTSETNASASLQSAATSASTATTKASEAATSARDAAASKEAAKSSE 250

Qy       66 GIVDAAGKADAKEGKLDATG----AEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAA 121
              :|:  | :      | |    |  |:  ||     : ||     |    | |||  :|
Db      251 --TNASSSASSAASSATAAGNSAKAAKTSETNARS--SETAAGQSASAAAGSKTAAASSA 306

Qy      122 SAATGNAAIGDVVNGDVAKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGAEGT 181
            |||: :|              |  :||     |  :    :|  |  | |:     :
Db      307 SAASTSA-------------GQASASATAAGKSAESAASSASTATTKAGEATEQASAAA 352

Qy      182 TNADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKA-AD 240
             :| | |    ||      : :: | ||| :| |      :   |:||   :|   | :
Db      353 RSASAAKTSETNA-KASETSAESSKTAAASSASSAASSASS-ASASKDEATRQASAAKSS 410

Qy      241 RTNPIDAAIGGAGDNDAAA------------AFATMKKDDQIAAAMVLRGMAKDGQFALK 288
             |     |   ||  |||              |    |: : ||:|:           ||:
Db      411 ATTASTKATEAAGSATAAAQSKSTAESAATRAETAAKRAEDIASAV-----------ALE 459

Qy      289 DAAAAHEGTVKNAVDIIKAAAEAASAASAATGSAAIGDVVNG--------NGATA--KGG 338
            ||:   :|-|:     : :|  : |   |||  |       |          |||    ||
Db      460 DASTTKKGIVQ-----LSSATNSTSETLAATPKAVKSAYDNAEKRLQKDQNGADIPDKGC 514

Qy      339 DAKSVNGIAK 348
                 ::| ::|
Db      515 FLNNINAVSK 524


RESULT  11
Y747_MYCTU
ID   Y747_MYCTU     STANDARD;      PRT;   801 AA.
AC   O53810;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   HYPOTHETICAL PE-PGRS FAMILY PROTEIN RV0747 PRECURSOR.
GN   RV0747 OR MT0772.5 OR MTV041.21.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC       SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL021958; CAA17514.1; -.
DR   EMBL; AE006968; AAK45011.1; -.
DR   TIGR; MT0772.5; -.
DR   TubercuList; Rv0747; -.
DR   InterPro; IPR000084; PE.
DR   Pfam; PF00934; PE; 1.
KW   Hypothetical protein; Repeat; Signal; Complete proteome.
FT   SIGNAL        1     30       POTENTIAL.
FT   CHAIN        31    801       HYPOTHETICAL PE-PGRS FAMILY PROTEIN
FT                                RV0747.
FT   CONFLICT    188    188       G -> S (IN REF. 2).
FT   CONFLICT    225    225       R -> G (IN REF. 2).
FT   CONFLICT    227    227       R -> G (IN REF. 2).
FT   CONFLICT    295    295       K -> R (IN REF. 2).
FT   CONFLICT    300    300       S -> G (IN REF. 2).
FT   CONFLICT    338    338       T -> I (IN REF. 2).
FT   CONFLICT    377    377       A -> P (IN REF. 2).
```

```
FT   CONFLICT    577    577       T -> A (IN REF. 2).
FT   CONFLICT    580    595       MISSING (IN REF. 2).
SQ   SEQUENCE   801 AA;  65407 MW;  EA54C9BF45A00F41 CRC64;


  Query Match             10.7%;  Score 183.5;  DB 1;  Length 801;
  Best Local Similarity   27.0%;  Pred. No. 0.014;
  Matches  107;  Conservative   29;  Mismatches  162;  Indels   99;  Gaps   19;

Qy      19 AKAAEEAASAAS----------AATGNAAIGD----------------VVKNSGAAAKG 51
           | || |||: :          ||||   ||:              :: | ||   |
Db      86 AYAAAEAAAVSPLLAPINAQFVAATGRPLIGNGANGAPGTGANGGPGGWLIGNGGAGGSG 145

Qy      52 --GEAASVNGIAKGIKGIVDAAG-KADAKEGKLDATGAEGTTNVNAGKLFVKRAADDG-- 106
             |  |  || | |: |   | |   |   | | |    ||| ||   ||  |
Db     146 APGAGAGGNGGAGGLFGSGGAGGASTDVAGGAGGAGGAGG----NAGMLF--GAAGVGGV 199

Qy     107 ---------GDADDAGKAAAAVAASAATGNAAIGDVVNGDVAKAKGGDA---ASVNGIAK 154
                  | | || |      |    |:   |::  |  |   ||:|   |: :| | |
Db     200 GGFSNGGATGGAGGAGGAGGLFGAGRERGSGGSGNLTGG--AGGAGGNAGTLATGDGGAG 257

Qy     155 GIKGIVDAAEKADAKEGKLNAAGAEGTTNADAGKLFVK----NAGNVGGEAGD--AGKAA 208
           | |          :: |    || |    |||  |        || :    ||  || |
Db     258 GTGGA--------SRSGGFGGAGGAG---GDAGMFFGSGGSGGAGGISKSVGDSAAGGAG 306

Qy     209 AAVAAV-------SGEQILKAIVHAAKDGGEKQGKKAADRTNPIDAAIGGAGDNDAAAAF 261
             |   :       :|          |  ||   |   |  |  :   ||  |:  :
Db     307 GAPGLIGNGGNGGNG---------GASTGGGDGGPGGAGGTGVL---IGNGGNGGSGGTG 354

Qy     262 ATMKKDDQIAAAMVLRGMAKDGQFALKDAAAAHE------GTVKNAVDIIKAAAEAASAA 315
           ||: |        || |:  ||  |    :  |        | :    :       : |
Db     355 ATLGKAGIGGTGGVLLGL--DGFTAPASTSPLHTLQQDVINMVNDPFQTLTGRPLIGNGA 412

Qy     316 SAATGSAAIGDV---VNGNGATAKGGDAKSVNGIAKG 349
           :   |: | |     : |||     |    ||| | |
Db     413 NGTPGTGADGGAGGWLFGNGGNGGQGTIGGVNGGAGG 449


RESULT  12
YQ34_MYCTU
ID   YQ34_MYCTU     STANDARD;      PRT;   778 AA.
AC   P71933;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   HYPOTHETICAL PE-PGRS FAMILY PROTEIN RV2634C.
GN   RV2634C OR MT2712 OR MTCY441.04C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC       SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z80225; CAB02341.1; -.
DR   EMBL; AE007103; AAK47026.1; ALT_INIT.
DR   HSSP; P04002; 1ATF.
DR   TIGR; MT2712; -.
DR   TubercuList; Rv2634c; -.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR000084; PE.
DR   Pfam; PF00934; PE; 1.
DR   ProDom; PD001223; PE; 1.
KW   Hypothetical protein; Complete proteome.
FT   CONFLICT     51     51       V -> L (IN REF. 2).
FT   CONFLICT     63     63       Q -> H (IN REF. 2).
FT   CONFLICT    274    274       A -> T (IN REF. 2).
SQ   SEQUENCE   778 AA;  63131 MW;  DAB20FE58E4999E7 CRC64;


  Query Match             10.7%;  Score 182;  DB 1;  Length 778;
  Best Local Similarity   26.6%;  Pred. No. 0.016;
  Matches   89;  Conservative   24;  Mismatches  130;  Indels   92;  Gaps   17;

Qy      29 ASAATGNAAIGD-VVKNSGAAAKGGEAASVNGIAKGIKGIVDAAGKAD----AKEGKLDA 83
           | || ||   |  :: : ||  :|| |  | |:  |: |     | |:       |
Db     314 AGAAGGNGGAGGMLLGDGGAGGQGGGPA--VAGVLGGMPGAGGNGGNANWFGSGGAGGQGG 371

Qy      84 TGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGDVVNGDVAKAKG 143
           ||  ||  || | :       :| |             ::  ||   |   ||    | |
Db     372 TGLAGTNGVNPGSIANPNTGANGTD------------NSGNGNQTGG---NGGPGPAGG 415

Qy     144 -GDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGAEGTTNADAGKLFVKNAGNVGGEAG 202
            |:|  | |  :|  |     |  |  :|     || ||   | |      || :|   |
Db     416 VGEAGGVGG--QGGLG-----ESLDGNDGTGGKGGAGGTAGTDGGAGGAGGAGGIGETDG 468

Qy     203 DAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAADRTNPIDAAIGGAG-------DN 255
            ||  |                        ||| |   |  |  :|  :|||       :
Db     469 SAGGVAT-------------------GGE--GGIGA--TGGVDGGVGGAGGKGGQGHNT 504

Qy     256 DAAAAFATMKKDDQIAAAMVLRGMAKDGQFALKDAAAAHEGTVKNAVDIIKAAAEAASAA 315
           ||      |         |:  ||  ||  ||  : ||            :
Db     505 GVGDAFG----GD--------GGIGGDGNGAL-GAAGGNGGT------------GGAGG 538

Qy     316 SAATGSAAIGD----VVNGNGATAKGGDAKSVNGI 346
           :   |   ||:       | | |  || |    |:
Db     539 NGGRGGMLIGNGGAGGAGGTGGTGGGGAAGFAGGV 573


RESULT  13
TOLA_ECOLI
ID   TOLA_ECOLI     STANDARD;      PRT;   421 AA.
AC   P19934;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   TOLA PROTEIN.
GN   TOLA OR CIM OR EXCC OR LKY OR B0739.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JM105;
RX   MEDLINE=90078104; PubMed=2687247;
RA   Levengood S.K., Webster R.E.;
RT   "Nucleotide sequences of the tolA and tolB genes and localization of
RT   their products, components of a multistep translocation system in
RT   Escherichia coli.";
RL   J. Bacteriol. 171:6600-6609(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97061202; PubMed=8905232;
RA   Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA   Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA   Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA   Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA   Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA   Yano M., Horiuchi T.;
RT   "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 12.7-28.0 min region on the linkage map.";
RL   DNA Res. 3:137-155(1996).
RN   [4]
RP   DOMAINS.
RX   MEDLINE=91296736; PubMed=2068069;
RA   Levengood S.K., Beyer W.F. Jr., Webster R.E.;
RT   "TolA: a membrane protein involved in colicin uptake contains an
RT   extended helical region.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:5939-5943(1991).
RN   [5]
RP   INTERACTION WITH PORINS.
RX   MEDLINE=97133271; PubMed=8978668;
RA   Derouiche R., Gavioli M., Benedetti H., Prilipov A., Lazdunski C.,
RA   Lloubes R.;
RT   "TolA central domain interacts with Escherichia coli porins.";
RL   EMBO J. 15:6408-6415(1996).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS) OF 298-421.
RX   MEDLINE=99332679; PubMed=10404600;
RA   Lubkowski J., Hennecke F., Plueckthun A., Wlodawer A.;
RT   "Filamentous phage infection: crystal structure of g3p in complex
RT   with its coreceptor, the C-terminal domain of TolA.";
RL   Structure 7:711-722(1999).
CC   -!- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF GROUP A
CC       COLICINS (COLICINS A, E1, E2, E3, AND K). NECESSARY FOR THE
CC       COLICINS TO REACH THEIR RESPECTIVE TARGETS AFTER INITIAL
CC       BINDING TO THE BACTERIA. ALSO INVOLVED IN THE TRANSLOCATION
CC       OF BACTERIOPHAGE DNA.
CC   -!- SUBUNIT: INTERACTS, VIA DOMAIN II, WITH PORINS OMPC, OMPC, PHOE
CC       AND LAMB.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. INNER MEMBRANE.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M28232; AAA24683.1; -.
DR   EMBL; AE000177; AAC73833.1; -.
DR   EMBL; D90713; BAA35405.1; -.
DR   PIR; JV0057; JV0057.
DR   PDB; 1TOL; 20-MAY-99.
DR   EcoGene; EG11007; tolA.
KW   Transport; Protein transport; Bacteriocin transport; Transmembrane;
KW   Repeat; Inner membrane; 3D-structure; Complete proteome.
FT   DOMAIN        1     13       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     14     34       POTENTIAL.
FT   DOMAIN       35    421       PERIPLASMIC (POTENTIAL).
FT   DOMAIN       48    310       DOMAIN II (ALPHA-HELICAL).
FT   DOMAIN      311    421       DOMAIN III (FUNCTIONAL).
FT   DOMAIN      224    278       10 X TANDEM REPEATS OF [ED]-K(1,2)-
FT                                A(2,4).
SQ   SEQUENCE   421 AA;  43156 MW;  8B2F52B4B97C655E CRC64;


  Query Match             10.6%;  Score 181.5;  DB 1;  Length 421;
  Best Local Similarity   27.4%;  Pred. No. 0.0099;
  Matches  101;  Conservative   44;  Mismatches  141;  Indels   83;  Gaps   15;

Qy      18 MAKAAEEAASAASAATGNAAIGDVV----------KNSGAAAKGGEAASVNGIAKGIKGI 67
           |:|| |:       |   :| :  ::          :|  |:| ||  :|:: :      :
Db       1 MSKATEQNDKLKRAIIISAVLHVILFAALIWSSFDENIEASAGGGGGSSIDAVMVDSGAV 60

Qy      68 VD----------AAGKADAKEGKLDATGAEGTTNVNAG------KLFVKR--AADDGGDA 109
           |:          :| ::| :    :   ||      |       :|  :|  | :    |
Db      61 VEQYKRMQSQESSAKRSDEQRKMKEQQAAEELREKQAAEQERLKQLEKERLAAQEQKKQA 120

Qy     110 DDAGKAA------AAVAASAATGNAAIGDVVNGDVAKAKG-GDAASVNGIAKGIKGIVDA 162
           ::| | |      |  ||: |  :           ||||   || :    ||  |   ||
Db     121 EEAAKQAELKQKQAEEAAAKAAAD-----------AKAKAEADAKAAEEAAK--KAAADA 167

Qy     163 AEKADAKEGKLNAAGAEGTTNADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQILKA 222
           :||:|:  |   | ||    |:|    :|             | | |  | : |   ||
Db     168 KKKAEAEAAK---AAAEAQKKAEAAAAALK------------KKAEAAEAAAAEARKKA 211

Qy     223 IVHAAKDGGEKQGKKAADRTNPIDAAIGGAGDNDAAAAFATM-KKDDQIAAAMVLRGMAK 281
           ||:    :  |||         ||   | |  |||  |   ||  : |||      |
Db     212 ATEAAEKAKAEAEKKA--------AAEKAAADKKAAAEKAAADKKAAEKAAA---EKAAA 260

Qy     282 DGQFALKDAAAAHEGTVKNAVDIIKAAAEAASAASAATGSAAI-GDVVNGNGATAKGGDA 340
           | : | : |||  :           |||||  |:|| ||  :  | |:: :|  |   || |
Db     261 DKKAAAEKAAADKKAAA------AKAAAEKAAAAKAAAEADDIFGELSSGKNAPKTGGGA 314

Qy     341 KSVNGIAKG 349
           |  |    |
Db     315 KGNNASPAG 323


RESULT  14
ELS_RAT
ID   ELS_RAT         STANDARD;      PRT;   864 AA.
AC   Q99372;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   ELASTIN PRECURSOR (TROPOELASTIN) (FRAGMENT).
GN   ELN.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91104868; PubMed=1702999;
RA   Pierce R.A., Deak S.B., Stolle C.A., Boyd C.D.;
RT   "Heterogeneity of rat tropoelastin mRNA revealed by cDNA cloning.";
RL   Biochemistry 29:9677-9683(1990).
RN   [2]
RP   SEQUENCE OF 781-864 FROM N.A.
RX   MEDLINE=88330868; PubMed=2971041;
RA   Deak S.B., Pierce R.A., Belsky S.A., Riley D.J., Boyd C.D.;
RT   "Rat tropoelastin is synthesized from a 3.5-kilobase mRNA.";
RL   J. Biol. Chem. 263:13504-13507(1988).
```

```
RN   [3]
RP   SEQUENCE OF 264-533 AND 558-864 FROM N.A.
RX   MEDLINE=92241859; PubMed=1572637;
RA   Pierce R.A., Alatawi A., Deak S.B., Boyd C.D.;
RT   "Elements of the rat tropoelastin gene associated with alternative
RT   splicing.";
RL   Genomics 12:651-658(1992).
CC   -!- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
CC       NUCHAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
CC   -!- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
CC       INTO AN EXTENSIBLE 3D NETWORK.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
CC   -!- ALTERNATIVE PRODUCTS: THREE DIFFERENT MRNAS HAVE BEEN FOUND THAT
CC       ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M60647; AAA42269.1; -.
DR   EMBL; J04035; AAA42268.1; -.
DR   EMBL; M86372; AAA42271.1; -.
DR   EMBL; M86355; AAA42271.1; JOINED.
DR   EMBL; M86363; AAA42271.1; JOINED.
DR   EMBL; M86364; AAA42271.1; JOINED.
DR   EMBL; M86366; AAA42271.1; JOINED.
DR   EMBL; M86371; AAA42271.1; JOINED.
DR   EMBL; M86376; AAA42272.1; -.
DR   EMBL; M86373; AAA42272.1; JOINED.
DR   EMBL; M86375; AAA42272.1; JOINED.
DR   HSSP; P04002; 1WFA.
KW   Structural protein; Connective tissue; Repeat; Signal;
KW   Alternative splicing.
FT   NON_TER       1      1
FT   SIGNAL       <1     21       BY SIMILARITY.
FT   CHAIN         22    864       ELASTIN.
FT   VARIANT      263    307       MISSING (IN CERTAIN CLONES).
FT   VARIANT      308    308       MISSING (IN CERTAIN CLONES).
FT   VARIANT      809    823       MISSING (IN CERTAIN CLONES).
SQ   SEQUENCE   864 AA;  72786 MW;  456894BB09E79FD4 CRC64;


  Query Match           10.5%;  Score 179;  DB 1;  Length 864;
  Best Local Similarity   26.3%;  Pred. No. 0.024;
  Matches  113;  Conservative   20;  Mismatches  192;  Indels  104;  Gaps   14;

Qy      19 AKAAEEAASAA--SAATGNAAIGDVVKNSGAAAKG-----------------GEAASVN 58
           | ||: || ||   || |    |  |:  ||   |                 |    |
Db     346 AAAAKAAAKAAKYGAAGGLVPGGPGVRVPGAGIPGVGIPGVGGIPGVGGIPGVGGIPGVG 405

Qy      59 GIAKGIKGIVD--------AAGKADAKEGKLDATGAEG--TTNVNAGKL---FVKRAADD 105
           |   |  |||         || || ||  |  | |  |  |  | ||      |   |
Db     406 GPGIGGPGIVGGPGAVSPAAAAKAAAKAAKYGARGGVGIPTYGVGAGGFPGYGVGAGAGL 465

Qy     106 GGDADDAGKAAAAVAASAATGNA-AIGDVVNGDVAKA---------KGGDAASVNGIAKG 155
           || :  |  |||| ||    | |  :| :| | |  |          |    :: |   |
Db     466 GGASQAAAAAAAAKAAKYGAGGAGTLGGLVPGAVPGALPGAVPGALPGAVPGALPGAVPG 525

Qy     156 IKG-------------IVDAAEKADAKEGKL------------------NAAGAEGTTN 183
           : |              || || || |:                    |  | |
Db     526 VPGTGGVPGAGTPAAAAAAAAAKAAAKAGQYGLGPGVGGVPGGVGVGGLPGGVGPGGVTG 585

Qy     184 ADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKD-GGEKQGKKAADRT 242
           |       | | |: | ||  ||| :|   :   ||   ||   |    |
Db     586 IGTG----PGTGLVPGDLGGAGTPAAAKSA--AKAAAKAQYRAAAGLGAGVPGLGVGAGV 639

Qy     243 NPIDAAIGGAGDNDAAAAFATMKKDDQIAAAMVLR----------------GMAKDGQF 285
               | || |       |        :||:   :                 |:   | |
```

```
Db     640 PGFGAGAGGFGAGAGVPGFGAGAVPGSLAASKAAKYGAAGGLGGPGGLGGPGGLGGPGGF 699

Qy     286 ALKDAAAAHEGTVKNAVDIIKAAAEAASAAS-----AATGSAAIGDVVNGNGATAKGGDA 340
                 | |        |||:||: |:      | |  | |    | ||   |
Db     700 GGPGGLGGVPGGVAGGAPAAAAAAKAAAKAAQYGLGGAGGLGAGGLGAGGLGAGGLGAGG 759

Qy     341 KSVNGIAKG 349
              |:  |
Db     760 LGAGGLGAG 768


RESULT  15
ELS_MOUSE
ID   ELS_MOUSE      STANDARD;      PRT;   860 AA.
AC   P54320;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   ELASTIN PRECURSOR (TROPOELASTIN).
GN   ELN.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=Lung;
RX   MEDLINE=95130069; PubMed=7829060;
RA   Wydner K.S., Sechler J.L., Boyd C.D., Passmore H.C.;
RT   "Use of an intron polymorphism to localize the tropoelastin gene to
RT   mouse chromosome 5 in a region of linkage conservation with human
RT   chromosome 7.";
RL   Genomics 23:125-131(1994).
CC   -!- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
CC       NUCHAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
CC   -!- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
CC       INTO AN EXTENSIBLE 3D NETWORK.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
CC   -!- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U08210; AAA80155.1; -.
DR   HSSP; P04002; 1WFA.
DR   MGD; MGI:95317; Eln.
KW   Structural protein; Repeat; Signal; Connective tissue.
FT   SIGNAL        1     27       POTENTIAL.
FT   CHAIN        28    860       ELASTIN.
SQ   SEQUENCE   860 AA;  71955 MW;  0C0BE5AAE1EDD7F1 CRC64;


  Query Match           10.4%;  Score 177;  DB 1;  Length 860;
  Best Local Similarity   27.4%;  Pred. No. 0.03;
  Matches  113;  Conservative   20;  Mismatches  185;  Indels   94;  Gaps   14;

Qy      19 AKAAEEAASAA--SAATGNAAIGDVVKNSGAAAKG------GEAASVNGIAKGIKGIVD 69
           | ||: || ||   || |    |  |:  ||   |      |    | |  |  |||
Db     360 AAAAKAAAKAAKYGAAGGLVPGGPGVRLPGAGIPGVGGIPGVGGIPGVGGPGIGGPGIVG 419

Qy      70 --------AAGKADAKEGKLDATGAEG--TTNVNAGKL---FVKRAADDGGDADDAGKAA 116
                  || || ||  |  | |  |  |  | ||      |   |  || :  |  ||
Db     420 GPGAVSPAAAAKAAAKAAKYGARGGVGIPTYGVGACGFPGYGVGAGAGLGGASPAAAAAA 479

Qy     117 AAVAASAATGNAAIGDVVNGDVAKAKGGDAASVNGIAKGIKG-------IVDAAEKADAK 169
           |  |   | |  |:| :| | |  |  |   :| | | |: |          || || ||
Db     480 AKAAKYGAGGAGALGGLVPGAVPGALPGAVPAVPG-AGGVPGAGTPAAAAAAAAAKAAAK 538
```

```
Qy    170 E-------------------------GKLNAAGAEGTTNADAGKLFVKNAGNVGGEAGDA 204
                                    |     |  | |   ||       | :||    |
Db    539 AGLGPGVGGVPGGVGVGGIPGGVGVGGVPGGVGPGGVTGIGAG------PGGLGGAGSPA 592

Qy    205 GKAAAAVAAVSGEQILKAIVHA------AKDGGEKQGKKAADRTNPIDAAIGGAGDNDAA 258
            :|| ||   :    | : |      |  |    |  |        | : | |
Db    593 AAKSAAKAAAKAQYRAAAGLGAGVPGFGAGAGVPGFGAGAGVPGFGAGAGVPGFGAGAGV 652

Qy    259 AAFATMKKDDQIAAAMVLR-----------------GMAKDGQFALKDAAAAHEGTVKNA 301
            |        :||:   :                  |:   |        |   | |  |
Db    653 PGFGAGAVPGSLAASKAAKYGAAGGLGGPGGLGGPGGLGGPGGL----GGAGVPGRVAGA 708

Qy    302 VDIIKAAAEAASAASAAT----GSAAIGDVVNGNGATAKGGDAKSVNGIAKG 349
              ||| | :|| ||     |:  :|    | ||   |       |:  |
Db    709 APPAAAAAAAAKAAAKAAQYGLGGAGGLG--AGGLGAGGLGAGGLGAGGLGAG 758


Search completed: January 10, 2002, 14:09:57
Job time: 326 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 10, 2002, 14:08:51 ; Search time 18.38 Seconds
(without alignments)
86.433 Million cell updates/sec

Title: US-09-445-803-14
Perfect score: 235
Sequence: 1 PLTNPIDAAIGGSADRNAEA.....DQIAAAMVLRGMAKDGQFAL 47

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 155098 seqs, 33800819 residues

Total number of hits satisfying chosen parameters: 155098

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_New:*
1: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep:*
2: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep:*
3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep:*
4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep:*
5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep:*
6: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:*
7: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	54	23.0	198	5	US-09-708-427-66722	Sequence 66722, A
2	54	23.0	312	5	US-09-708-427-66721	Sequence 66721, A
3	54	23.0	394	5	US-09-708-427-66720	Sequence 66720, A
4	54	23.0	604	5	US-09-708-427-28541	Sequence 28541, A
5	54	23.0	620	5	US-09-708-427-28540	Sequence 28540, A
6	54	23.0	686	5	US-09-708-427-28539	Sequence 28539, A
7	53	22.6	359	5	US-09-897-516-5275	Sequence 5275, Ap
8	53	22.6	526	5	US-09-708-427-27506	Sequence 27506, A
9	53	22.6	544	5	US-09-708-427-27505	Sequence 27505, A
10	53	22.6	727	5	US-09-708-427-27504	Sequence 27504, A
11	52	22.1	1480	5	US-09-708-427-11856	Sequence 11856, A
12	52	22.1	1535	5	US-09-708-427-11855	Sequence 11855, A
13	52	22.1	1548	5	US-09-708-427-11854	Sequence 11854, A
14	51.5	21.9	114	5	US-09-154-707A-319	Sequence 319, App
15	51.5	21.9	114	5	US-09-966-262-319	Sequence 319, App
16	51	21.7	92	5	US-09-897-516-7117	Sequence 7117, Ap
17	51	21.7	127	5	US-09-611-526-2565	Sequence 2565, Ap
18	51	21.7	167	5	US-09-708-427-62406	Sequence 62406, A
19	51	21.7	201	5	US-09-708-427-62405	Sequence 62405, A
20	51	21.7	649	5	US-09-708-427-12799	Sequence 12799, A
21	50	21.3	174	5	US-09-708-427-65815	Sequence 65815, A
22	50	21.3	174	5	US-09-708-427-83240	Sequence 83240, A
23	50	21.3	423	5	US-09-708-427-70072	Sequence 70072, A
24	50	21.3	498	5	US-09-815-242-11777	Sequence 11777, A
25	50	21.3	585	5	US-09-708-427-25137	Sequence 25137, A
26	50	21.3	625	5	US-09-708-427-29600	Sequence 29600, A
27	50	21.3	626	5	US-09-708-427-25136	Sequence 25136, A
28	50	21.3	630	5	US-09-708-427-25135	Sequence 25135, A
29	50	21.3	648	5	US-09-468-646A-30	Sequence 30, Appl
30	50	21.3	648	5	US-09-468-646B-30	Sequence 30, Appl
31	50	21.3	675	5	US-09-708-427-29599	Sequence 29599, A
32	50	21.3	695	5	US-09-708-427-29598	Sequence 29598, A
33	49.5	21.1	177	5	US-09-708-427-76073	Sequence 76073, A
34	49.5	21.1	183	5	US-09-708-427-76072	Sequence 76072, A
35	49.5	21.1	193	5	US-09-708-427-76071	Sequence 76071, A
36	49.5	21.1	344	5	US-09-620-394B-966	Sequence 966, App
37	49.5	21.1	375	5	US-09-620-394B-965	Sequence 965, App
38	49.5	21.1	393	5	US-09-620-394B-964	Sequence 964, App
39	49	20.9	127	5	US-09-708-427-82550	Sequence 82550, A
40	49	20.9	139	5	US-09-708-427-62795	Sequence 62795, A
41	49	20.9	225	5	US-09-708-427-62794	Sequence 62794, A
42	49	20.9	262	5	US-09-708-427-62793	Sequence 62793, A
43	49	20.9	362	5	US-09-708-427-18950	Sequence 18950, A
44	49	20.9	371	5	US-09-708-427-18949	Sequence 18949, A
45	49	20.9	409	5	US-09-708-427-18948	Sequence 18948, A

ALIGNMENTS

```
RESULT 1
US-09-708-427-66722
; Sequence 66722, Application US/09708427
; GENERAL INFORMATION:
;  APPLICANT: N. ALEXANDROV et al.
;  TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
;  TITLE OF INVENTION:  THEREBY
;  FILE REFERENCE: 2750-1243P
;  CURRENT APPLICATION NUMBER: US/09/708,427
;  CURRENT FILING DATE:  2000-11-09
;  NUMBER OF SEQ ID NOS: 85364
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 66722
;   LENGTH: 198
;   TYPE: PRT
;   ORGANISM: Zea mays subsp. mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1..198
;   OTHER INFORMATION: Xaa is any amino acid
;   NAME/KEY: misc_feature
;   LOCATION: 1..198
;   OTHER INFORMATION: Ceres Seq. ID 1931147
US-09-708-427-66722


  Query Match             23.0%;  Score 54;  DB 5;  Length 198;
  Best Local Similarity   27.9%;  Pred. No. 6.8;
  Matches   12;  Conservative   11;  Mismatches   20;  Indels    0;  Gaps    0;

Qy       5 PIDAAIGGSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
            |: : :  |   ||:|| : :  |  |  : || ::|| :  :
Db     122 PLQSCMNHSCRPNAKAFKRDEDKDGHAVIIALRPISKDEEITI 164


RESULT 2
US-09-708-427-66721
; Sequence 66721, Application US/09708427
; GENERAL INFORMATION:
;  APPLICANT: N. ALEXANDROV et al.
;  TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
;  TITLE OF INVENTION:  THEREBY
;  FILE REFERENCE: 2750-1243P
;  CURRENT APPLICATION NUMBER: US/09/708,427
;  CURRENT FILING DATE:  2000-11-09
;  NUMBER OF SEQ ID NOS: 85364
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 66721
;   LENGTH: 312
```

```
;   TYPE: PRT
;   ORGANISM: Zea mays subsp. mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1..312
;   OTHER INFORMATION: Xaa is any amino acid
;   NAME/KEY: misc_feature
;   LOCATION: 1..312
;   OTHER INFORMATION: Ceres Seq. ID 1931146
US-09-708-427-66721


  Query Match             23.0%;  Score 54;  DB 5;  Length 312;
  Best Local Similarity   27.9%;  Pred. No. 12;
  Matches   12;  Conservative   11;  Mismatches   20;  Indels    0;  Gaps    0;

Qy       5 PIDAAIGGSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
           |: : :  |   ||:|| : :  |  |  : || ::|| :  :
Db     236 PLQSCMNHSCRPNAKAFKRDEDKDGHAVIIALRPISKDEEITI 278


RESULT   3
US-09-708-427-66720
; Sequence 66720, Application US/09708427
; GENERAL INFORMATION:
;  APPLICANT: N. ALEXANDROV et al.
;  TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
;  TITLE OF INVENTION:  THEREBY
;  FILE REFERENCE: 2750-1243P
;  CURRENT APPLICATION NUMBER: US/09/708,427
;  CURRENT FILING DATE:  2000-11-09
;  NUMBER OF SEQ ID NOS: 85364
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 66720
;   LENGTH: 394
;   TYPE: PRT
;   ORGANISM: Zea mays subsp. mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1..394
;   OTHER INFORMATION: Xaa is any amino acid
;   NAME/KEY: misc_feature
;   LOCATION: 1..394
;   OTHER INFORMATION: Ceres Seq. ID 1931145
US-09-708-427-66720


  Query Match             23.0%;  Score 54;  DB 5;  Length 394;
  Best Local Similarity   27.9%;  Pred. No. 16;
  Matches   12;  Conservative   11;  Mismatches   20;  Indels    0;  Gaps    0;

Qy       5 PIDAAIGGSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
           |: : :  |   ||:|| : :  |  |  : || ::|| :  :
Db     318 PLQSCMNHSCRPNAKAFKRDEDKDGHAVIIALRPISKDEEITI 360


RESULT   4
US-09-708-427-28541
; Sequence 28541, Application US/09708427
; GENERAL INFORMATION:
;  APPLICANT: N. ALEXANDROV et al.
;  TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
;  TITLE OF INVENTION:  THEREBY
;  FILE REFERENCE: 2750-1243P
;  CURRENT APPLICATION NUMBER: US/09/708,427
;  CURRENT FILING DATE:  2000-11-09
;  NUMBER OF SEQ ID NOS: 85364
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28541
;   LENGTH: 604
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1..604
;   OTHER INFORMATION: Xaa is any amino acid
;   NAME/KEY: misc_feature
;   LOCATION: 1..604
;   OTHER INFORMATION: Ceres Seq. ID 1823502
US-09-708-427-28541


  Query Match             23.0%;  Score 54;  DB 5;  Length 604;
  Best Local Similarity   36.6%;  Pred. No. 27;
  Matches   15;  Conservative    5;  Mismatches   19;  Indels    2;  Gaps    1;

Qy       1 PLTNPIDAAIG--GSADRNAEAFDKMKKDDQIAAAMVLRGM 39
           || | :    |  |:|    : ||:|   | || | ||  :
Db      25 PLANTLVNVYGKCGAASHALQVFDEMPHRDHIAWASVLTAL 65


RESULT   5
US-09-708-427-28540
; Sequence 28540, Application US/09708427
; GENERAL INFORMATION:
;  APPLICANT: N. ALEXANDROV et al.
;  TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
;  TITLE OF INVENTION:  THEREBY
;  FILE REFERENCE: 2750-1243P
;  CURRENT APPLICATION NUMBER: US/09/708,427
;  CURRENT FILING DATE:  2000-11-09
;  NUMBER OF SEQ ID NOS: 85364
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28540
;   LENGTH: 620
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1..620
;   OTHER INFORMATION: Xaa is any amino acid
;   NAME/KEY: misc_feature
;   LOCATION: 1..620
;   OTHER INFORMATION: Ceres Seq. ID 1823501
US-09-708-427-28540


  Query Match             23.0%;  Score 54;  DB 5;  Length 620;
  Best Local Similarity   36.6%;  Pred. No. 27;
  Matches   15;  Conservative    5;  Mismatches   19;  Indels    2;  Gaps    1;

Qy       1 PLTNPIDAAIG--GSADRNAEAFDKMKKDDQIAAAMVLRGM 39
           || | :    |  |:|    : ||:|   | || | ||  :
Db      41 PLANTLVNVYGKCGAASHALQVFDEMPHRDHIAWASVLTAL 81


RESULT   6
US-09-708-427-28539
; Sequence 28539, Application US/09708427
; GENERAL INFORMATION:
;  APPLICANT: N. ALEXANDROV et al.
;  TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
;  TITLE OF INVENTION:  THEREBY
;  FILE REFERENCE: 2750-1243P
;  CURRENT APPLICATION NUMBER: US/09/708,427
;  CURRENT FILING DATE:  2000-11-09
;  NUMBER OF SEQ ID NOS: 85364
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28539
;   LENGTH: 686
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1..686
;   OTHER INFORMATION: Xaa is any amino acid
```

```
;   NAME/KEY: misc_feature
;   LOCATION: 1..686
;   OTHER INFORMATION: Ceres Seq. ID 1823500
US-09-708-427-28539


  Query Match             23.0%;  Score 54;  DB 5;  Length 686;
  Best Local Similarity   36.6%;  Pred. No. 31;
  Matches   15;  Conservative    5;  Mismatches   19;  Indels    2;  Gaps    1;

Qy       1 PLTNPIDAAIG--GSADRNAEAFDKMKKDDQIAAAMVLRGM 39
           || | :    |  |:|    : ||:|   | || | ||  :
Db     107 PLANTLVNVYGKCGAASHALQVFDEMPHRDHIAWASVLTAL 147


RESULT   7
US-09-897-516-5275
; Sequence 5275, Application US/09897516
; GENERAL INFORMATION:
;  APPLICANT: Corbin, David R.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Huesing, Joseph E.
;  APPLICANT:  Krasomil-Osterfeld, Karina C.
;  APPLICANT:  Malvar, Thomas M.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Spiridonov, Sergei
;  TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
;  FILE REFERENCE: 38-21(51847)B
;  CURRENT APPLICATION NUMBER: US/09/897,516
;  CURRENT FILING DATE:  2001-06-29
;  PRIOR APPLICATION NUMBER: US 60/215, 161
;  PRIOR FILING DATE: 2000-06-30
;  NUMBER OF SEQ ID NOS: 8409
; SEQ ID NO 5275
;   LENGTH: 359
;   TYPE: PRT
;   ORGANISM: Xenorhabdus sp.
US-09-897-516-5275


  Query Match             22.6%;  Score 53;  DB 5;  Length 359;
  Best Local Similarity   39.3%;  Pred. No. 19;
  Matches   11;  Conservative    5;  Mismatches   12;  Indels    0;  Gaps    0;

Qy      16 RNAEAFDKMKKDDQIAAAMVLRGMAKDG 43
           :||: ::  || |      |:||  |||
Db     219 KNAKLYEAAKKADGTDGGYVIRGEDKDG 246


RESULT   8
US-09-708-427-27506
; Sequence 27506, Application US/09708427
; GENERAL INFORMATION:
;  APPLICANT: N. ALEXANDROV et al.
;  TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
;  TITLE OF INVENTION:  THEREBY
;  FILE REFERENCE: 2750-1243P
;  CURRENT APPLICATION NUMBER: US/09/708,427
;  CURRENT FILING DATE:  2000-11-09
;  NUMBER OF SEQ ID NOS: 85364
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27506
;   LENGTH: 526
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1..526
;   OTHER INFORMATION: Xaa is any amino acid
;   NAME/KEY: misc_feature
;   LOCATION: 1..526
;   OTHER INFORMATION: Ceres Seq. ID 1821102
US-09-708-427-27506


  Query Match             22.6%;  Score 53;  DB 5;  Length 526;
  Best Local Similarity   29.7%;  Pred. No. 30;
  Matches   11;  Conservative   11;  Mismatches   11;  Indels    4;  Gaps    1;

Qy      12 GSADRNAEAFDKMKKD----DQIAAAMVLRGMAKDGQ 44
           |: | :   :|:|::|    | :|   :: |: |||:
Db     111 GNLDASLRVWDEMRRDEIKPDVMAYGTLVVGLCKDGR 147


RESULT   9
US-09-708-427-27505
; Sequence 27505, Application US/09708427
; GENERAL INFORMATION:
;  APPLICANT: N. ALEXANDROV et al.
;  TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
;  TITLE OF INVENTION:  THEREBY
;  FILE REFERENCE: 2750-1243P
;  CURRENT APPLICATION NUMBER: US/09/708,427
;  CURRENT FILING DATE:  2000-11-09
;  NUMBER OF SEQ ID NOS: 85364
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27505
;   LENGTH: 544
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1..544
;   OTHER INFORMATION: Xaa is any amino acid
;   NAME/KEY: misc_feature
;   LOCATION: 1..544
;   OTHER INFORMATION: Ceres Seq. ID 1821101
US-09-708-427-27505


  Query Match             22.6%;  Score 53;  DB 5;  Length 544;
  Best Local Similarity   29.7%;  Pred. No. 32;
  Matches   11;  Conservative   11;  Mismatches   11;  Indels    4;  Gaps    1;

Qy      12 GSADRNAEAFDKMKKD----DQIAAAMVLRGMAKDGQ 44
           |: | :   :|:|::|    | :|   :: |: |||:
Db     129 GNLDASLRVWDEMRRDEIKPDVMAYGTLVVGLCKDGR 165


RESULT  10
US-09-708-427-27504
; Sequence 27504, Application US/09708427
; GENERAL INFORMATION:
;  APPLICANT: N. ALEXANDROV et al.
;  TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
;  TITLE OF INVENTION:  THEREBY
;  FILE REFERENCE: 2750-1243P
;  CURRENT APPLICATION NUMBER: US/09/708,427
;  CURRENT FILING DATE:  2000-11-09
;  NUMBER OF SEQ ID NOS: 85364
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27504
;   LENGTH: 727
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1..727
;   OTHER INFORMATION: Xaa is any amino acid
;   NAME/KEY: misc_feature
;   LOCATION: 1..727
;   OTHER INFORMATION: Ceres Seq. ID 1821100
US-09-708-427-27504
```

Query Match 22.6%; Score 53; DB 5; Length 727;
Best Local Similarity 29.7%; Pred. No. 45;
Matches 11; Conservative 11; Mismatches 11; Indels 4; Gaps 1;

```
Qy      12 GSADRNAEAFDKMKKD----DQIAAAMVLRGMAKDGQ 44
           |: | :   :|:|::|    | :|   :: |: |||:
Db     312 GNLDASLRVWDEMRRDEIKPDVMAYGTLVVGLCKDGR 348
```

RESULT 11
US-09-708-427-11856
; Sequence 11856, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: THEREBY
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11856
; LENGTH: 1480
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..1480
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..1480
; OTHER INFORMATION: Ceres Seq. ID 1823914
US-09-708-427-11856

Query Match 22.1%; Score 52; DB 5; Length 1480;
Best Local Similarity 38.5%; Pred. No. 1.4e+02;
Matches 10; Conservative 7; Mismatches 9; Indels 0; Gaps 0;

```
Qy      19 EAFDKMKKDDQIAAAMVLRGMAKDGQ 44
           : |||| : | :|   |: | |::|:
Db     109 KVFDKMPEKDLVAWNSVINGFAENGK 134
```

RESULT 12
US-09-708-427-11855
; Sequence 11855, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: THEREBY
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11855
; LENGTH: 1535
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..1535
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..1535
; OTHER INFORMATION: Ceres Seq. ID 1823913
US-09-708-427-11855

Query Match 22.1%; Score 52; DB 5; Length 1535;
Best Local Similarity 38.5%; Pred. No. 1.5e+02;
Matches 10; Conservative 7; Mismatches 9; Indels 0; Gaps 0;

```
Qy      19 EAFDKMKKDDQIAAAMVLRGMAKDGQ 44
           : |||| : | :|   |: | |::|:
Db     164 KVFDKMPEKDLVAWNSVINGFAENGK 189
```

RESULT 13
US-09-708-427-11854
; Sequence 11854, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
; TITLE OF INVENTION: THEREBY
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11854
; LENGTH: 1548
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..1548
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..1548
; OTHER INFORMATION: Ceres Seq. ID 1823912
US-09-708-427-11854

Query Match 22.1%; Score 52; DB 5; Length 1548;
Best Local Similarity 38.5%; Pred. No. 1.5e+02;
Matches 10; Conservative 7; Mismatches 9; Indels 0; Gaps 0;

```
Qy      19 EAFDKMKKDDQIAAAMVLRGMAKDGQ 44
           : |||| : | :|   |: | |::|:
Db     177 KVFDKMPEKDLVAWNSVINGFAENGK 202
```

RESULT 14
US-09-154-707A-319
; Sequence 319, Application US/09154707A
; GENERAL INFORMATION:
; APPLICANT: Young et al.
; TITLE OF INVENTION: 87 Human Secreted Proteins
; FILE REFERENCE: PZ004P1
; CURRENT APPLICATION NUMBER: US/09/154,707A
; CURRENT FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: PCT/US98/05311
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: US 60/041,277
; PRIOR FILING DATE: 1997-03-21
; PRIOR APPLICATION NUMBER: US 60/042,344
; PRIOR FILING DATE: 1997-03-21
; PRIOR APPLICATION NUMBER: US 60/041,276
; PRIOR FILING DATE: 1997-03-21
; PRIOR APPLICATION NUMBER: US 60/041,281
; PRIOR FILING DATE: 1997-03-21
; PRIOR APPLICATION NUMBER: US 60/048,094
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,350
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,188
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,135
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/050,937
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,187
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,099

```
;  PRIOR FILING DATE: 1997-05-30
;  PRIOR APPLICATION NUMBER: US 60/048,352
;  PRIOR FILING DATE: 1997-05-30
;  PRIOR APPLICATION NUMBER: US 60/048,186
;  PRIOR FILING DATE: 1997-05-30
;  PRIOR APPLICATION NUMBER: US 60/048,069
;  PRIOR FILING DATE: 1997-05-30
;  PRIOR APPLICATION NUMBER: US 60/048,095
;  PRIOR FILING DATE: 1997-05-30
;  PRIOR APPLICATION NUMBER: US 60/048,131
;  PRIOR FILING DATE: 1997-05-30
;  PRIOR APPLICATION NUMBER: US 60/048,096
;  PRIOR FILING DATE: 1997-05-30
;  PRIOR APPLICATION NUMBER: US 60/048,355
;  PRIOR FILING DATE: 1997-05-30
;  PRIOR APPLICATION NUMBER: US 60/048,160
;  PRIOR FILING DATE: 1997-05-30
;  PRIOR APPLICATION NUMBER: US 60/048,351
;  PRIOR FILING DATE: 1997-05-30
;  PRIOR APPLICATION NUMBER: US 60/048,154
;  PRIOR FILING DATE: 1997-05-30
;  PRIOR APPLICATION NUMBER: US 60/054,804
;  PRIOR FILING DATE: 1997-08-05
;  PRIOR APPLICATION NUMBER: US 60/056,370
;  PRIOR FILING DATE: 1997-08-19
;  PRIOR APPLICATION NUMBER: US 60/060,862
;  PRIOR FILING DATE: 1997-10-02
;  NUMBER OF SEQ ID NOS: 343
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 319
;   LENGTH: 114
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (2)
;   OTHER INFORMATION: Xaa equals any of the L-amino acids commonly found in naturally o
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (114)
;   OTHER INFORMATION: Xaa equals any of the L-amino acids commonly found in naturally o
US-09-154-707A-319


  Query Match           21.9%;  Score 51.5;  DB 5;  Length 114;
  Best Local Similarity  34.1%;  Pred. No. 7.5;
  Matches   14;  Conservative    7;  Mismatches   17;  Indels    3;  Gaps    1;

Qy       5 PIDAAIGGSADRNAEAFDKMKKDDQIAAAM---VLRGMAKD 42
           |::|| |  ||    | ::  : :  | ||    | | |:|
Db       9 PVEAAAGAGADGREPASERASRAEPPAVAMGQNDLMGTAED 49


RESULT  15
US-09-966-262-319
; Sequence 319, Application US/09966262
; GENERAL INFORMATION:
;  APPLICANT: Young et al.
;  TITLE OF INVENTION: 87 Human Secreted Proteins
;  FILE REFERENCE: PZ004P1
;  CURRENT APPLICATION NUMBER: US/09/966,262
;  CURRENT FILING DATE:  2001-10-01
;  PRIOR APPLICATION NUMBER: US 09/154,707
;  PRIOR FILING DATE: 1998-09-17
;  PRIOR APPLICATION NUMBER: PCT/US98/05311
;  PRIOR FILING DATE: 1998-03-19
;  PRIOR APPLICATION NUMBER: US 60/041,277
;  PRIOR FILING DATE: 1997-03-21
;  PRIOR APPLICATION NUMBER: US 60/042,344
;  PRIOR FILING DATE: 1997-03-21
;  PRIOR APPLICATION NUMBER: US 60/041,276
;  PRIOR FILING DATE: 1997-03-21
;  PRIOR APPLICATION NUMBER: US 60/041,281
;  PRIOR FILING DATE: 1997-03-21
;  PRIOR APPLICATION NUMBER: US 60/048,094
;  PRIOR FILING DATE: 1997-05-30
;  PRIOR APPLICATION NUMBER: US 60/048,350
;  PRIOR FILING DATE: 1997-05-30
;  PRIOR APPLICATION NUMBER: US 60/048,188
;  PRIOR FILING DATE: 1997-05-30
;  PRIOR APPLICATION NUMBER: US 60/048,135
;  PRIOR FILING DATE: 1997-05-30
;  PRIOR APPLICATION NUMBER: US 60/050,937
;  PRIOR FILING DATE: 1997-05-30
;  PRIOR APPLICATION NUMBER: US 60/048,187
;  PRIOR FILING DATE: 1997-05-30
;  PRIOR APPLICATION NUMBER: US 60/048,099
;  PRIOR FILING DATE: 1997-05-30
;  PRIOR APPLICATION NUMBER: US 60/048,352
;  PRIOR FILING DATE: 1997-05-30
;  PRIOR APPLICATION NUMBER: US 60/048,186
;  PRIOR FILING DATE: 1997-05-30
;  PRIOR APPLICATION NUMBER: US 60/048,069
;  PRIOR FILING DATE: 1997-05-30
;  PRIOR APPLICATION NUMBER: US 60/048,095
;  PRIOR FILING DATE: 1997-05-30
;  PRIOR APPLICATION NUMBER: US 60/048,131
;  PRIOR FILING DATE: 1997-05-30
;  PRIOR APPLICATION NUMBER: US 60/048,096
;  PRIOR FILING DATE: 1997-05-30
;  PRIOR APPLICATION NUMBER: US 60/048,355
;  PRIOR FILING DATE: 1997-05-30
;  PRIOR APPLICATION NUMBER: US 60/048,160
;  PRIOR FILING DATE: 1997-05-30
;  PRIOR APPLICATION NUMBER: US 60/048,351
;  PRIOR FILING DATE: 1997-05-30
;  PRIOR APPLICATION NUMBER: US 60/048,154
;  PRIOR FILING DATE: 1997-05-30
;  PRIOR APPLICATION NUMBER: US 60/054,804
;  PRIOR FILING DATE: 1997-08-05
;  PRIOR APPLICATION NUMBER: US 60/056,370
;  PRIOR FILING DATE: 1997-08-19
;  PRIOR APPLICATION NUMBER: US 60/060,862
;  PRIOR FILING DATE: 1997-10-02
;  NUMBER OF SEQ ID NOS: 343
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 319
;   LENGTH: 114
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (2)
;   OTHER INFORMATION: Xaa equals any of the L-amino acids commonly found in naturall
;   OTHER INFORMATION: proteins
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (114)
;   OTHER INFORMATION: Xaa equals any of the L-amino acids commonly found in naturall
;   OTHER INFORMATION: proteins
US-09-966-262-319


  Query Match           21.9%;  Score 51.5;  DB 5;  Length 114;
  Best Local Similarity  34.1%;  Pred. No. 7.5;
  Matches   14;  Conservative    7;  Mismatches   17;  Indels    3;  Gaps    1;

Qy       5 PIDAAIGGSADRNAEAFDKMKKDDQIAAAM---VLRGMAKD 42
           |::|| |  ||    | ::  : :  | ||    | | |:|
Db       9 PVEAAAGAGADGREPASERASRAEPPAVAMGQNDLMGTAED 49


Search completed: January 10, 2002, 14:08:52
Job time: 321 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 10, 2002, 14:03:56 ; Search time 19.72 Seconds
(without alignments)
53.634 Million cell updates/sec

Title: US-09-445-803-14
Perfect score: 235
Sequence: 1 PLTNPIDAAIGGSADRNAEA.....DQIAAAMVLRGMAKDGQFAL 47

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	50.5	21.5	1253	1	US-08-252-966B-12	Sequence 12, Appl
2	50.5	21.5	1261	1	US-08-252-966B-18	Sequence 18, Appl
3	49	20.9	553	2	US-08-793-229-36	Sequence 36, Appl
4	49	20.9	553	3	US-09-285-957-36	Sequence 36, Appl
5	48.5	20.6	344	1	US-07-941-523-24	Sequence 24, Appl
6	48	20.4	170	4	US-09-117-257-50	Sequence 50, Appl
7	47	20.0	388	2	US-08-282-197C-56	Sequence 56, Appl
8	47	20.0	486	2	US-08-942-423-3	Sequence 3, Appli
9	47	20.0	524	2	US-08-615-942A-2	Sequence 2, Appli
10	47	20.0	803	4	US-09-063-035-2	Sequence 2, Appli
11	47	20.0	1455	2	US-08-726-012B-2	Sequence 2, Appli
12	47	20.0	1618	1	US-07-853-913-4	Sequence 4, Appli
13	47	20.0	3033	1	US-07-925-695-8	Sequence 8, Appli
14	47	20.0	3033	1	US-07-925-695-9	Sequence 9, Appli
15	46.5	19.8	1338	2	US-08-728-470-9	Sequence 9, Appli
16	46.5	19.8	1338	4	US-08-719-641-9	Sequence 9, Appli
17	46.5	19.8	1529	2	US-08-728-470-10	Sequence 10, Appl
18	46.5	19.8	1529	4	US-08-719-641-10	Sequence 10, Appl
19	46.5	19.8	1599	2	US-08-617-697-9	Sequence 9, Appli
20	46.5	19.8	1600	2	US-08-617-697-10	Sequence 10, Appl
21	46	19.6	381	1	US-08-482-385A-6	Sequence 6, Appli
22	46	19.6	4551	3	US-09-320-878-1	Sequence 1, Appli
23	46	19.6	4613	4	US-09-105-537-31	Sequence 31, Appl
24	46	19.6	11877	4	US-09-105-537-6	Sequence 6, Appli
25	45.5	19.4	382	3	US-09-320-878-7	Sequence 7, Appli
26	45.5	19.4	402	4	US-09-105-537-22	Sequence 22, Appl
27	45.5	19.4	645	4	US-09-202-712-2	Sequence 2, Appli
28	45.5	19.4	3782	4	US-09-105-537-4	Sequence 4, Appli
29	45	19.1	338	1	US-08-210-394-1	Sequence 1, Appli
30	45	19.1	506	1	US-08-369-780-2	Sequence 2, Appli
31	45	19.1	506	1	US-08-475-682-2	Sequence 2, Appli
32	45	19.1	506	1	US-08-780-833-2	Sequence 2, Appli
33	45	19.1	506	1	US-08-636-036-2	Sequence 2, Appli
34	45	19.1	506	3	US-08-918-509-2	Sequence 2, Appli
35	45	19.1	506	3	US-09-108-262-2	Sequence 2, Appli
36	45	19.1	578	1	US-08-458-120-2	Sequence 2, Appli
37	45	19.1	578	2	US-08-867-970-2	Sequence 2, Appli
38	45	19.1	578	4	US-09-326-217-2	Sequence 2, Appli
39	44.5	18.9	32	3	US-08-484-322-50	Sequence 50, Appl
40	44.5	18.9	318	2	US-08-671-947-2	Sequence 2, Appli
41	44.5	18.9	861	4	US-08-960-048-12	Sequence 12, Appl
42	44	18.7	35	4	US-09-082-279B-939	Sequence 939, App
43	44	18.7	35	4	US-09-082-279B-940	Sequence 940, App
44	44	18.7	486	2	US-08-942-423-2	Sequence 2, Appli
45	44	18.7	706	1	US-08-339-152A-29	Sequence 29, Appl

ALIGNMENTS

RESULT 1
US-08-252-966B-12
; Sequence 12, Application US/08252966B
; Patent No. 5624818
; GENERAL INFORMATION:
; APPLICANT: Eisenman, Robert N.
; APPLICANT: Hurlin, Peter J.
; APPLICANT: Ayer, Donald E.
; TITLE OF INVENTION: Regulatory Proteins that Dimerize with
; TITLE OF INVENTION: Mad or Max
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen, O'Connor, Johnson, and KindnessPLLC
; STREET: 1420 Fifth Ave., Suite 2800
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/252,966B
; FILING DATE: 01-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Shelton, Dennis K.
; REGISTRATION NUMBER: 26,997.
; REFERENCE/DOCKET NUMBER: FHCR17694
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 682-8100
; TELEFAX: (206) 224-0779
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1253 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: translation of msina cDNA; see Figure 23
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Mus musculus
US-08-252-966B-12

Query Match 21.5%; Score 50.5; DB 1; Length 1253;
Best Local Similarity 26.7%; Pred. No. 58;

```
  Matches   12;  Conservative   10;  Mismatches   10;  Indels   13;  Gaps    1;

Qy       6 IDAAIGGS------------ADRNAEAFDKMKKDDQIAAAMVLR 37
           :|  :||:             ||: |:  | ::|:  ||  :||:
Db     672 LDNTLGGTSEVIHRKALQRIYADKAADIIDGLRKNPSIAVPIVLK 716


RESULT   2
US-08-252-966B-18
; Sequence 18, Application US/08252966B
; Patent No. 5624818
;  GENERAL INFORMATION:
;    APPLICANT:  Eisenman, Robert N.
;    APPLICANT:  Hurlin, Peter J.
;    APPLICANT:  Ayer, Donald E.
;    TITLE OF INVENTION:  Regulatory Proteins that Dimerize with
;    TITLE OF INVENTION:  Mad or Max
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen, O'Connor, Johnson, and KindnessPLLC
;      STREET:  1420 Fifth Ave., Suite 2800
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/252,966B
;      FILING DATE:  01-JUN-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Shelton, Dennis K.
;      REGISTRATION NUMBER:  26,997.
;      REFERENCE/DOCKET NUMBER:  FHCR17694
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 682-8100
;      TELEFAX:  (206) 224-0779
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1261 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;      DESCRIPTION:  translation of msina9 cDNA; see Figure 29A, B, C, D
;    HYPOTHETICAL:  YES
;    ORIGINAL SOURCE:
;      ORGANISM:  Mus musculus
US-08-252-966B-18


  Query Match             21.5%;  Score 50.5;  DB 1;  Length 1261;
  Best Local Similarity   26.7%;  Pred. No. 58;
  Matches   12;  Conservative   10;  Mismatches   10;  Indels   13;  Gaps    1;

Qy       6 IDAAIGGS------------ADRNAEAFDKMKKDDQIAAAMVLR 37
           :|  :||:             ||: |:  | ::|:  ||  :||:
Db     672 LDNTLGGTSEVIHRKALQRIYADKAADIIDGLRKNPSIAVPIVLK 716


RESULT   3
US-08-793-229-36
; Sequence 36, Application US/08793229
; Patent No. 5891703
;  GENERAL INFORMATION:
;    APPLICANT:  VAN DER LAAN, Jan Metske
;    APPLICANT:  RIEMENS, Adriana Marina
;    APPLICANT:  QUAX, Wilhelmus Johannes
;    TITLE OF INVENTION:  Mutated Penicillin G Acylase Genes
;    NUMBER OF SEQUENCES:  36
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  McDonnell Boehnen Hulbert & Berghoff
;      STREET:  300 South Wacker Drive
;      CITY:  Chicago
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/793,229
;      FILING DATE:  23-APR-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP95/03249
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      REFERENCE/DOCKET NUMBER:  97075
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312)913-0001
;      TELEFAX:  (312)913-0002
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  553 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-793-229-36


  Query Match             20.9%;  Score 49;  DB 2;  Length 553;
  Best Local Similarity   43.3%;  Pred. No. 34;
  Matches   13;  Conservative    3;  Mismatches    8;  Indels    6;  Gaps    1;

Qy       8 AAIGGSADR------NAEAFDKMKKDDQIA 31
           | : |||||        ||:||:  ||  |
Db     255 AFLWGSADRVKEIDNRIEAYDKLTADDMWA 284


RESULT   4
US-09-285-957-36
; Sequence 36, Application US/09285957
; Patent No. 6033823
;  GENERAL INFORMATION:
;    APPLICANT:  VAN DER LAAN, Jan Metske
;    APPLICANT:  RIEMENS, Adriana Marina
;    APPLICANT:  QUAX, Wilhelmus Johannes
;    TITLE OF INVENTION:  Mutated Penicillin G Acylase Genes
;    NUMBER OF SEQUENCES:  36
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  McDonnell Boehnen Hulbert & Berghoff
;      STREET:  300 South Wacker Drive
;      CITY:  Chicago
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/285,957
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/793,229
;      FILING DATE:
```

```
;     ATTORNEY/AGENT INFORMATION:
;       REFERENCE/DOCKET NUMBER:  97075
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (312)913-0001
;       TELEFAX:  (312)913-0002
;   INFORMATION FOR SEQ ID NO:  36:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  553 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-09-285-957-36


  Query Match             20.9%;  Score 49;  DB 3;  Length 553;
  Best Local Similarity   43.3%;  Pred. No. 34;
  Matches   13;  Conservative    3;  Mismatches    8;  Indels    6;  Gaps    1;

Qy      8 AAIGGSADR------NAEAFDKMKKDDQIA 31
          | : |||||        ||:||:  ||  |
Db    255 AFLWGSADRVKEIDNRIEAYDKLTADDMWA 284


RESULT   5
US-07-941-523-24
; Sequence 24, Application US/07941523
; Patent No. 5571718
;  GENERAL INFORMATION:
;     APPLICANT:  Dunn, John J
;     APPLICANT:  Barbour, Alan G
;     TITLE OF INVENTION:  Cloning and Expression of Borrelia
;     TITLE OF INVENTION:  Lipoproteins
;     NUMBER OF SEQUENCES:  24
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;       STREET:  Two Militia Drive
;       CITY:  Lexington
;       STATE:  Massachusetts
;       COUNTRY:  U.S.A.
;       ZIP:  01730
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/941,523
;       FILING DATE:  19920908
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Granahan, Patricia
;       REGISTRATION NUMBER:  32,227
;       REFERENCE/DOCKET NUMBER:  BNL90-01A
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 861-6240
;       TELEFAX:  (617) 861-9540
;   INFORMATION FOR SEQ ID NO:  24:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  344 amino acids
;       TYPE:  AMINO ACID
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-07-941-523-24


  Query Match             20.6%;  Score 48.5;  DB 1;  Length 344;
  Best Local Similarity   26.4%;  Pred. No. 22;
  Matches   14;  Conservative    6;  Mismatches   20;  Indels   13;  Gaps    1;

Qy      7 DAAIGGSADRNAEAFDKMK------------KDDQIAAAMVLRGMAKDGQFA 46
          || :  :|| |  |    |             |   :|  : || : | |: |
Db    229 DAQLAAAADGNTSAISFAKGGSDAHLAGANTPKAAAVAGGIALRSLVKTGKLA 281


RESULT   6
US-09-117-257-50
; Sequence 50, Application US/09117257
; Patent No. 6214355
; GENERAL INFORMATION:
;  APPLICANT: Hook, Magnus
;  APPLICANT: Guo, Betty
;  APPLICANT: Hanson, Mark
;  TITLE OF INVENTION: DbpA AND DbpB COMPOSITIONS AND METHODS OF USE
;  FILE REFERENCE: 4210.000500
;  CURRENT APPLICATION NUMBER: US/09/117,257
;  CURRENT FILING DATE: 1998-07-22
;  EARLIER APPLICATION NUMBER: PCT/US96/17081
;  EARLIER FILING DATE: 1996-10-22
;  EARLIER APPLICATION NUMBER: 08/589,711
;  EARLIER FILING DATE: 1996-01-22
;  EARLIER APPLICATION NUMBER: 08/427,023
;  EARLIER FILING DATE: 1995-04-24
;  NUMBER OF SEQ ID NOS: 66
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 50
;   LENGTH: 170
;   TYPE: PRT
;   ORGANISM: Borrelia afzelii
US-09-117-257-50


  Query Match             20.4%;  Score 48;  DB 4;  Length 170;
  Best Local Similarity   30.6%;  Pred. No. 11;
  Matches   11;  Conservative    9;  Mismatches   14;  Indels    2;  Gaps    1;

Qy      2 LTNPIDAAIGGSADR--NAEAFDKMKKDDQIAAAMV 35
          :|| || ||  : |   | :|| : :   ::| : :
Db     36 ITNEIDKAIKAAKDAGVNTDAFTETQTGGKVAGSQI 71


RESULT   7
US-08-282-197C-56
; Sequence 56, Application US/08282197C
; Patent No. 5871730
;  GENERAL INFORMATION:
;     APPLICANT:  Brzezinski, Ryszard
;     APPLICANT:  Dery, Claude V
;     APPLICANT:  Beaulieu, Carole
;     TITLE OF INVENTION:  Thermostable Xylanase DNA, Protein and
;     TITLE OF INVENTION:  Methods of Use
;     NUMBER OF SEQUENCES:  67
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Sterne, Kessler, Goldstein & Fox P.L.L.C.
;       STREET:  1100 New York Ave., NW
;       CITY:  Washington
;       STATE:  DC
;       COUNTRY:  USA
;       ZIP:  20005
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/282,197C
;       FILING DATE:  29-JUL-1994
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Cimbala, Michele A
;       REGISTRATION NUMBER:  33,851
;       REFERENCE/DOCKET NUMBER:  1050.0410000
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  202-371-2600
;       TELEFAX:  202-371-2540
```

```
;  INFORMATION FOR SEQ ID NO:  56:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  388 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  both
US-08-282-197C-56


  Query Match             20.0%;  Score 47;  DB 2;  Length 388;
  Best Local Similarity   37.0%;  Pred. No. 44;
  Matches   10;  Conservative    7;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      14 ADRNAEAFDKMKKDDQIAAAMVLRGMA 40
           | |  | || :|::  | :|:|  |::
Db     320 AKRYQELFDALKENKDIVSAVVFWGIS 346


RESULT   8
[redacted]8-942-423-3
;  Sequence 3, Application US/08942423
;  Patent No. 5891673
;  GENERAL INFORMATION:
;    APPLICANT:  Hashimoto, Yasuhiro
;    APPLICANT:  Takemoto, Yoshihiro
;    TITLE OF INVENTION:  Lck Binding Protein
;    NUMBER OF SEQUENCES:  68
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Syntex (U.S.A.) Inc.
;      STREET:  3401 Hillview Ave.
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  94303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/942,423
;      FILING DATE:  01-OCT-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/362,715
;      FILING DATE:  23-DEC-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Peries, Rohan
;      REGISTRATION NUMBER:  35,752
;      REFERENCE/DOCKET NUMBER:  28260
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 852-1698
;      TELEFAX:  (415) 496-3529
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  486 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:  N-terminal
;    ORIGINAL SOURCE:
;      ORGANISM:  LCK BINDING PROTEIN
US-08-942-423-3


  Query Match             20.0%;  Score 47;  DB 2;  Length 486;
  Best Local Similarity   33.3%;  Pred. No. 58;
  Matches   13;  Conservative    8;  Mismatches   14;  Indels    4;  Gaps    1;

Qy       3 TNPIDAAIGGSADRNAE----AFDKMKKDDQIAAAMVLR 37
            |  ||:||  |:    |:    | :| |::::  |  | |
Db     225 TTPIEAASSGARGLKAKFESMAEEKRKREEEEKAQQVAR 263


RESULT   9
US-08-615-942A-2
;  Sequence 2, Application US/08615942A
;  Patent No. 5863532
;  GENERAL INFORMATION:
;    APPLICANT:  JOLINDA A. TRAUGH
;    APPLICANT:  REGINA D. ROONEY
;    APPLICANT:  ROLF JAKOBI
;    APPLICANT:  POLYGENA T. TUAZON
;    APPLICANT:  CHARNG-JUI CHEN
;    APPLICANT:  WILLIAM E. MEEK
;    APPLICANT:  EDWARD J. CARROLL, JR.
;    APPLICANT:  CURTIS A. MONNIG
;    TITLE OF INVENTION:  Compositions and Methods Comprising
;    TITLE OF INVENTION:  Cytostatic Protein Kinase
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Robbins, Berliner & Carson
;      STREET:  201 N. Figueroa St., Suite 500
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  90012
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/615,942A
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  NONE
;      FILING DATE:  NONE
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wong, Wean K.
;      REGISTRATION NUMBER:  33,561
;      REFERENCE/DOCKET NUMBER:  1279-203XX
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  213/977-1001
;      TELEFAX:  213/977-1003
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  524 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-615-942A-2


  Query Match             20.0%;  Score 47;  DB 2;  Length 524;
  Best Local Similarity   39.4%;  Pred. No. 64;
  Matches   13;  Conservative    6;  Mismatches    8;  Indels    6;  Gaps    2;

Qy       4 NPIDAAIGGS-ADRNAEAFDKMKK-----DDQI 30
             :|| | :| |  |  |:: || ||     |::|
Db     200 DPIPAPVGDSHVDSGAKSSDKQKKKTKMTDEEI 232


RESULT  10
US-09-063-035-2
;  Sequence 2, Application US/09063035
;  Patent No. 6160091
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  PEUKERT, Karen; HAENEL, Frank; and EILERS,
;    APPLICANT:  Martin
;    TITLE OF INVENTION:  Myc-binding zinc finger proteins,
;    TITLE OF INVENTION:  their preparation and their use
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Keil & Weinkauf
;      STREET:  1101 Connecticut Avenue
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.2 Mb storage
;      COMPUTER:  IBM AT-compatible, 80486 processor
;      OPERATING SYSTEM:  MS-DOS version 6.1
;      SOFTWARE:  WordPerfect version 8.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/063,035
;      FILING DATE:  21-APR-1998
;      CLASSIFICATION:  514
;   INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  803 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-09-063-035-2


  Query Match             20.0%;  Score 47;  DB 4;  Length 803;
  Best Local Similarity   34.2%;  Pred. No. 1.1e+02;
  Matches   13;  Conservative    9;  Mismatches   10;  Indels    6;  Gaps    2;

Qy       1 PLTNPIDAAIGGSADRNA-EAFDKMKKDDQIAAAMVLR 37
           | |:|     ||:|:  | |  ||  |::::| : : |
Db     117 PATSP-----GGNAEALATEGGDKRAKEEKVATSTLSR 149


RESULT  11
US-08-726-012B-2
; Sequence 2, Application US/08726012B
; Patent No. 5952190
;  GENERAL INFORMATION:
;    APPLICANT:  Hans Joenje, et al.
;    TITLE OF INVENTION:  cDNA FOR FANCONI ANEMIA COMPLEMENTATION GROUP A
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Klarquist Sparkman Campbell Leigh & Whinston, LLP
;      STREET:  One World Trade Center, Suite 1600, 121 S.W. Salmon Street
;      CITY:  Portland
;      STATE:  OR
;      COUNTRY:  USA
;      ZIP:  97204-2988
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Disk, 3.5-inch
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  MS DOS
;      SOFTWARE:  WordPerfect 5.1+, ASCII
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/726,012B
;      FILING DATE:  10/04/96
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Richard J. Polley
;      REGISTRATION NUMBER:  28,107
;      REFERENCE/DOCKET NUMBER:  3812-45520/RJP/DJE
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (503) 226-7391
;      TELEFAX:  (503) 228-9446
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1455
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  Linear
US-08-726-012B-2


  Query Match             20.0%;  Score 47;  DB 2;  Length 1455;
  Best Local Similarity   34.5%;  Pred. No. 2.3e+02;
  Matches   10;  Conservative    7;  Mismatches   12;  Indels    0;  Gaps    0;

Qy      18 AEAFDKMKKDDQIAAAMVLRGMAKDGQFA 46
           |:||:  : |  : | :|:|  | :|  |
Db     416 AQAFESCQLDSMVTAFLVVRQAALEGPSA 444


RESULT  12
US-07-853-913-4
; Sequence 4, Application US/07853913
; Patent No. 5338839
;  GENERAL INFORMATION:
;    APPLICANT:  McKay, Ronald D.G.
;    APPLICANT:  Lendahl, Urban
;    TITLE OF INVENTION:  Nestin Expression As An Indicator of
;    TITLE OF INVENTION:  Neuroepithelial Tumors
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/853,913
;      FILING DATE:  19920319
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/660,412
;      FILING DATE:  22-FEB-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/603,803
;      FILING DATE:  25-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/201,762
;      FILING DATE:  02-JUN-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/180,548
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Granahan, Patricia
;      REGISTRATION NUMBER:  32,227
;      REFERENCE/DOCKET NUMBER:  MIT-4641AAAA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-861-6240
;      TELEFAX:  617-861-9540
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1618 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-853-913-4


  Query Match             20.0%;  Score 47;  DB 1;  Length 1618;
  Best Local Similarity   34.5%;  Pred. No. 2.6e+02;
  Matches   10;  Conservative    5;  Mismatches   14;  Indels    0;  Gaps    0;
```

```
Qy       1 PLTNPIDAAIGGSADRNAEAFDKMKKDDQ 29
           |||       :    | | |||  ::|::|
Db     690 PLTKENQEPLRSLEDENKEAFRSLEKENQ 718


RESULT  13
US-07-925-695-8
; Sequence 8, Application US/07925695
; Patent No. 5428145
;  GENERAL INFORMATION:
;    APPLICANT:  OKAMOTO, Hiroaki
;    APPLICANT:  NAKAMURA, Tetsuo
;    TITLE OF INVENTION:  NON-A, NON-B HEPATITIS VIRUS GENOME,
;    TITLE OF INVENTION:  POLYNUCLEOTIDES, POLYPEPTIDES, ANTIGEN, ANTIBODY AND
;    TITLE OF INVENTION:  DETECTION SYSTEMS
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Beveridge, DeGrandi, Weilacher & Young
;      STREET:  1850 M Street, N.W., Suite 800
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  US
;      ZIP:  20036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/925,695
;      FILING DATE:  19920807
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 287402/91
;      FILING DATE:  09-AUG-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 360441/91
;      FILING DATE:  05-DEC-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Weilacher, Robert G.
;      REGISTRATION NUMBER:  20,531
;      REFERENCE/DOCKET NUMBER:  06/87-48009
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 659-2811
;      TELEFAX:  (202) 659-1462
;      TELEX:  WUI 64470
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3033 amino acids
;      TYPE:  AMINO ACID
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  linear
US-07-925-695-8


  Query Match             20.0%;  Score 47;  DB 1;  Length 3033;
  Best Local Similarity   22.2%;  Pred. No. 5.8e+02;
  Matches   18;  Conservative    7;  Mismatches   20;  Indels   36;  Gaps    2;

Qy       1 PLTNPIDAAIGGSAD-----RNAEAFDKMKKDDQ------------------------- 29
           | |  :|    |: |     |||:     :|||:
Db    1119 PGTKSLDPCTCGAVDLYLVTRNADVIPVRRKDDRRGALLSPRPLSTLKGSSGGPVLCSRG 1178

Qy      30 -----IAAAMVLRGMAKDGQF 45
                 ||:  ||:||   |
Db    1179 HAVGLFRAAVCARGVAKSIDF 1199


RESULT  14
US-07-925-695-9
; Sequence 9, Application US/07925695
; Patent No. 5428145
;  GENERAL INFORMATION:
;    APPLICANT:  OKAMOTO, Hiroaki
;    APPLICANT:  NAKAMURA, Tetsuo
;    TITLE OF INVENTION:  NON-A, NON-B HEPATITIS VIRUS GENOME,
;    TITLE OF INVENTION:  POLYNUCLEOTIDES, POLYPEPTIDES, ANTIGEN, ANTIBODY AND
;    TITLE OF INVENTION:  DETECTION SYSTEMS
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Beveridge, DeGrandi, Weilacher & Young
;      STREET:  1850 M Street, N.W., Suite 800
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  US
;      ZIP:  20036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/925,695
;      FILING DATE:  19920807
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 287402/91
;      FILING DATE:  09-AUG-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 360441/91
;      FILING DATE:  05-DEC-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Weilacher, Robert G.
;      REGISTRATION NUMBER:  20,531
;      REFERENCE/DOCKET NUMBER:  06/87-48009
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 659-2811
;      TELEFAX:  (202) 659-1462
;      TELEX:  WUI 64470
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3033 amino acids
;      TYPE:  AMINO ACID
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  linear
US-07-925-695-9


  Query Match             20.0%;  Score 47;  DB 1;  Length 3033;
  Best Local Similarity   22.2%;  Pred. No. 5.8e+02;
  Matches   18;  Conservative    7;  Mismatches   20;  Indels   36;  Gaps    2;

Qy       1 PLTNPIDAAIGGSAD-----RNAEAFDKMKKDDQ------------------------- 29
           | |  :|    |: |     |||:     :|||:
Db    1119 PGTKSLDPCTCGAVDLYLVTRNADVIPVRRKDDRRGALLSPRPLSTLKGSSGGPVLCSRG 1178

Qy      30 -----IAAAMVLRGMAKDGQF 45
                 ||:  ||:||   |
Db    1179 HAVGLFRAAVCARGVAKSIDF 1199


RESULT  15
US-08-728-470-9
; Sequence 9, Application US/08728470
; Patent No. 5928651
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 5928651-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
```

; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,470
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-633
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1338 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-728-470-9

Query Match 19.8%; Score 46.5; DB 2; Length 1338;
Best Local Similarity 46.2%; Pred. No. 2.5e+02;
Matches 12; Conservative 4; Mismatches 9; Indels 1; Gaps 1;

```
Qy       6 IDAAIGGSADRNAEAFDKMKKDDQIA 31
           |  | ||:||     |||: || :|:
Db     804 IGNASGGNADAKKVTFDKV-KDSKIS 828
```

Search completed: January 10, 2002, 14:03:58
Job time: 92 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 10, 2002, 14:05:14 ; Search time 40.15 Seconds
(without alignments)
86.711 Million cell updates/sec

Title: US-09-445-803-14
Perfect score: 235
Sequence: 1 PLTNPIDAAIGGSADRNAEA.....DQIAAAMVLRGMAKDGQFAL 47

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101:*
1: /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	235	100.0	47	20	AAW95613	Borrelia burgdorfe
2	176.5	75.1	349	20	AAW95612	Borrelia burgdorfe
3	176.5	75.1	349	21	AAB36281	B. garinii P7-1 pr
4	144.5	61.5	1328	20	AAY20088	B. burgdorferi ant
5	142	60.4	156	20	AAY20113	B. burgdorferi ant
6	142	60.4	533	20	AAY20112	B. burgdorferi ant
7	131.5	56.0	168	20	AAY20089	B. burgdorferi ant
8	130	55.3	356	18	AAW22676	Borrelia variable
9	117	49.8	26	21	AAB36273	B. burgdorferi var
10	117	49.8	26	21	AAB36280	B. burgdorferi var
11	55	23.4	593	22	AAG91309	C glutamicum prote
12	52.5	22.3	354	22	AAG91164	C glutamicum prote
13	51.5	21.9	245	20	AAY21894	PHA acetoacetyl-Co
14	51	21.7	127	21	AAY73333	HTRM clone 1760185
15	51	21.7	383	19	AAW48336	Brucella abortus 3
16	50.5	21.5	969	21	AAB01827	Haemophilus influe
17	50.5	21.5	975	21	AAB01826	Haemophilus influe
18	50.5	21.5	1026	22	AAB94315	Human protein sequ
19	50.5	21.5	1253	18	AAW10038	Mad binding protei
20	50.5	21.5	1261	18	AAW10040	Mad binding protei
21	49.5	21.1	143	21	AAG45798	Arabidopsis thalia
22	49.5	21.1	143	21	AAG45827	Arabidopsis thalia
23	49.5	21.1	177	21	AAG33768	Arabidopsis thalia
24	49.5	21.1	183	21	AAG33767	Arabidopsis thalia
25	49.5	21.1	185	21	AAG17255	Arabidopsis thalia
26	49.5	21.1	185	21	AAG45797	Arabidopsis thalia
27	49.5	21.1	185	21	AAG45826	Arabidopsis thalia
28	49.5	21.1	193	21	AAG33766	Arabidopsis thalia
29	49.5	21.1	206	21	AAG17254	Arabidopsis thalia
30	49.5	21.1	206	21	AAG22849	Arabidopsis thalia
31	49.5	21.1	206	21	AAG45796	Arabidopsis thalia
32	49.5	21.1	206	21	AAG45825	Arabidopsis thalia
33	49.5	21.1	257	21	AAG17253	Arabidopsis thalia
34	49.5	21.1	257	21	AAG22848	Arabidopsis thalia
35	49.5	21.1	287	21	AAG22847	Arabidopsis thalia
36	49.5	21.1	297	22	AAG81486	S. epidermidis ope
37	49.5	21.1	297	22	AAG81826	S. epidermidis ope
38	49.5	21.1	311	21	AAG04819	Arabidopsis thalia
39	49.5	21.1	317	21	AAG04818	Arabidopsis thalia
40	49.5	21.1	319	22	AAG82375	S. epidermidis ope
41	49.5	21.1	324	21	AAB15081	Arabidopsis NAC1 p
42	49.5	21.1	327	21	AAG04817	Arabidopsis thalia
43	49.5	21.1	344	21	AAG13282	Arabidopsis thalia
44	49.5	21.1	375	21	AAG13281	Arabidopsis thalia
45	49.5	21.1	393	21	AAG13280	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAW95613
ID AAW95613 standard; Protein; 47 AA.
XX
AC AAW95613;
XX
DT 08-JUN-1999 (first entry)
XX
DE Borrelia burgdorferi surface antigen P39.5 clone 14-1 polypeptide.
XX
KW Lyme disease; surface antigen; P39.5; diagnosis; prevention;
KW vaccine; antisense; therapy; treatment; primer; probe;
KW antibody; DNA.
XX
OS Borrelia burgdorferi.
XX
FH Key Location/Qualifiers
FT CDS 2..142
FT /*tag= a
FT /product= "clone 14-1 encoded protein"
XX
PN WO9900413-A1.
XX
PD 07-JAN-1999.
XX
PF 29-JUN-1998; 98WO-US13551.
XX
PR 30-JUN-1997; 97US-0051271.
XX
PA (TULA) TULANE EDUCATIONAL FUND.
XX
PI Philipp MT;
XX
DR WPI; 1999-095676/08.

DR N-PSDB; AAX07422.
XX
PT New nucleic acid encoding the P39.5 antigen of Borrelia burgdorferi
PT - and related vectors, transformants, antibodies and polypeptides,
PT for diagnosis, prevention and treatment of Lyme disease
XX
PS Claim 5; Page 60; 89pp; English.
XX
CC The sequence is that of a Borrelia burgdorferi surface antigen P39.5
CC clone 14-1 polypeptide. It can be used in the production of P39.5 or
CC fragments of it which may be used to raise antibodies to, and in the
CC development of vaccines against Lyme disease. The sequence can
CC also be used for making primers and probes for diagnosis, also in
CC DNA vaccines, as antisense therapeutics and for drug screening.
CC Antibodiescan be used as diagnostic (immunoassay) reagents, for
CC treating Lyme disease, for affinity purification, for drug
CC screening and to produce anti-idiotype antibodies (used in the
CC same way as P39.5 to induce an immune response).
XX
SQ Sequence 47 AA;

Query Match 100.0%; Score 235; DB 20; Length 47;
Best Local Similarity 100.0%; Pred. No. 2.4e-26;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PLTNPIDAAIGGSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 pltnpidaaiggsadrnaeafdkmkkddqiaaamvlrgmakdgqfal 47

RESULT 2
AAW95612
ID AAW95612 standard; Protein; 349 AA.
XX
AC AAW95612;
XX
DT 08-JUN-1999 (first entry)
XX
DE Borrelia burgdorferi surface antigen P39.5 clone 7-1 polypeptide.
XX
KW Lyme disease; surface antigen; P39.5; diagnosis; prevention;
KW vaccine; antisense; therapy; treatment; primer; probe;
KW antibody; DNA.
XX
OS Borrelia burgdorferi.
XX
PN WO9900413-A1.
XX
PD 07-JAN-1999.
XX
PF 29-JUN-1998; 98WO-US13551.
XX
PR 30-JUN-1997; 97US-0051271.
XX
PA (TULA) TULANE EDUCATIONAL FUND.
XX
PI Philipp MT;
XX
DR WPI; 1999-095676/08.
DR N-PSDB; AAX07411.
XX
PT New nucleic acid encoding the P39.5 antigen of Borrelia burgdorferi
PT - and related vectors, transformants, antibodies and polypeptides,
PT for diagnosis, prevention and treatment of Lyme disease
XX
PS Claim 5; Pages 54-55; 89pp; English.
XX
CC The sequence is that of a Borrelia burgdorferi surface antigen P39.5
CC clone 7-1 polypeptide. It can be used in the production of P39.5
CC or fragments of it which may be used to raise antibodies to, and
CC in the development of vaccines against Lyme disease. The sequence
CC can also be used for making primers and probes for diagnosis, also
CC in DNA vaccines, as antisense therapeutics and for drug screening.
CC Antibodies can be used as diagnostic (immunoassay) reagents, for
CC treating Lyme disease, for affinity purification, for drug
CC screening and to produce anti-idiotype antibodies (used in the
CC same way as P39.5 to induce an immune response).
XX
SQ Sequence 349 AA;

Query Match 75.1%; Score 176.5; DB 20; Length 349;
Best Local Similarity 82.6%; Pred. No. 5.5e-17;
Matches 38; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

Qy 3 TNPIDAAIGGSADRN-AEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
|||||||||||: | : | || |||||||||||||||||||||||
Db 242 tnpidaaiggagdndaaaafatmkkddqiaaamvlrgmakdgqfal 287

RESULT 3
AAB36281
ID AAB36281 standard; peptide; 349 AA.
XX
AC AAB36281;
XX
DT 21-FEB-2001 (first entry)
XX
DE B. garinii P7-1 protein.
XX
KW Variable surface antigen; invariable region; VlsE; Lyme disease;
KW Lyme borreliosis.
XX
OS Borrelia garinii.
XX
PN WO200065064-A1.
XX
PD 02-NOV-2000.
XX
PF 25-APR-2000; 2000WO-US11085.
XX
PR 28-APR-1999; 99US-0300971.
XX
PA (TULA) TULANE EDUCATIONAL FUND.
XX
PI Philipp MT, Liang FT;
XX
DR WPI; 2000-687350/67.
XX
PT Novel peptides comprising an invariable 26-amino acid long region
PT isolated from Borrelia burgdorferi (sensu lato), useful for rapid and
PT specific diagnosis of lyme disease -
XX
PS Example 2; Fig 2; 76pp; English.
XX
CC The present invention describes several peptides comprised of the
CC invariable regions IR1-IR6 of the B. burgdorferi variable surface antigen
CC (VlsE) variable domain. These peptides can be used in the specific
CC diagnosis, treatment and vaccination against B. burgdorferi, which causes
CC Lyme disease (also known as Lyme borreliosis) in humans, dogs, horses,
CC cows and other animals.
XX
SQ Sequence 349 AA;

Query Match 75.1%; Score 176.5; DB 21; Length 349;
Best Local Similarity 82.6%; Pred. No. 5.5e-17;
Matches 38; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

Qy 3 TNPIDAAIGGSADRN-AEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
|||||||||||: | : | || |||||||||||||||||||||||
Db 242 tnpidaaiggagdndaaaafatmkkddqiaaamvlrgmakdgqfal 287

RESULT 4

```
AAY20088
ID   AAY20088 standard; Protein; 1328 AA.
XX
AC   AAY20088;
XX
DT   19-JUL-1999  (first entry)
XX
DE   B. burgdorferi antigenic protein, f24-1.aa.
XX
KW   Antigenic protein; vaccine; Lyme disease; infection; detection.
XX
OS   Borrelia burgdorferi.
XX
PN   WO9859071-A1.
XX
PD   30-DEC-1998.
XX
PF   18-JUN-1998;   98WO-US12718.
XX
PR   03-SEP-1997;   97US-0057483.
PR   20-JUN-1997;   97US-0050359.
PR   22-JUL-1997;   97US-0053344.
PR   22-JUL-1997;   97US-0053377.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Choi GH,  Erwin AL,  Hanson MS,  Lathigra R;
XX
DR   WPI; 1999-189980/16.
DR   N-PSDB; AAX61785.
XX
PT   New isolated Borrelia burgdorferi nucleic acids  - used to develop
PT   products for the diagnosis, prevention and treatment of diseases
PT   caused by Borrelia, particularly Lyme disease
XX
PS   Claim 12; Page 192-193; 275pp; English.
XX
CC   This sequence represents a Borrelia burgdorferi (Bb) protein of the
CC   invention, which is suitable for use in a vaccine. The Bb polypeptides
CC   can be used in vaccines for eliciting protective antibodies to members of
CC   the Borrelia genus, particularly for the use against Lyme disease in
CC   humans and animals. They can be used for preventing or attenuating an
CC   infection caused by a member of the Borrelia genus. The products can also
CC   be used for detection of members of the Borrelia genus.
XX
SQ   Sequence   1328 AA;

  Query Match             61.5%;  Score 144.5;  DB 20;  Length 1328;
  Best Local Similarity   71.1%;  Pred. No. 9.8e-12;
  Matches   32;  Conservative    5;  Mismatches    7;  Indels    1;  Gaps    1;

Qy       4 NPIDAAIG-GSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
           ||| |||| |:|| |:  | |||||||||||: |||||||||:||:
Db     479 npiaaaigkgnaddgadfgdgmkkddqiaaaialrgmakdgkfav 523


RESULT   5
AAY20113
ID   AAY20113 standard; Protein; 156 AA.
XX
AC   AAY20113;
XX
DT   19-JUL-1999  (first entry)
XX
DE   B. burgdorferi antigenic protein, t49-2.aa.
XX
KW   Antigenic protein; vaccine; Lyme disease; infection; detection.
XX
OS   Borrelia burgdorferi.
XX
PN   WO9859071-A1.
XX
PD   30-DEC-1998.
XX
PF   18-JUN-1998;   98WO-US12718.
XX
PR   03-SEP-1997;   97US-0057483.
PR   20-JUN-1997;   97US-0050359.
PR   22-JUL-1997;   97US-0053344.
PR   22-JUL-1997;   97US-0053377.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Choi GH,  Erwin AL,  Hanson MS,  Lathigra R;
XX
DR   WPI; 1999-189980/16.
DR   N-PSDB; AAX61810.
XX
PT   New isolated Borrelia burgdorferi nucleic acids  - used to develop
PT   products for the diagnosis, prevention and treatment of diseases
PT   caused by Borrelia, particularly Lyme disease
XX
PS   Claim 12; Page 202; 275pp; English.
XX
CC   This sequence represents a Borrelia burgdorferi (Bb) protein of the
CC   invention, which is suitable for use.in a vaccine. The Bb polypeptides
CC   can be used in vaccines for eliciting protective antibodies to members of
CC   the Borrelia genus, particularly for the use against Lyme disease in
CC   humans and animals. They can be used for preventing or attenuating an
CC   infection caused by a member of the Borrelia genus. The products can also
CC   be used for detection of members of the Borrelia genus.
XX
SQ   Sequence   156 AA;

  Query Match             60.4%;  Score 142;  DB 20;  Length 156;
  Best Local Similarity   68.2%;  Pred. No. 1.7e-12;
  Matches   30;  Conservative    5;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       4 NPIDAAIGGSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
           ||| ||||   :  ||  |:||||||||||: |||||||||:||:
Db     100 npiaaaigkgnedgaefkdemkkddqiaaaialrgmakdgkfav 143


RESULT   6
AAY20112
ID   AAY20112 standard; Protein; 533 AA.
XX
AC   AAY20112;
XX
DT   19-JUL-1999  (first entry)
XX
DE   B. burgdorferi antigenic protein, f49-2.aa.
XX
KW   Antigenic protein; vaccine; Lyme disease; infection; detection.
XX
OS   Borrelia burgdorferi.
XX
PN   WO9859071-A1.
XX
PD   30-DEC-1998.
XX
PF   18-JUN-1998;   98WO-US12718.
XX
PR   03-SEP-1997;   97US-0057483.
PR   20-JUN-1997;   97US-0050359.
PR   22-JUL-1997;   97US-0053344.
PR   22-JUL-1997;   97US-0053377.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Choi GH,  Erwin AL,  Hanson MS,  Lathigra R;
```

XX
DR WPI; 1999-189980/16.
DR N-PSDB; AAX61809.
XX
PT New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the diagnosis, prevention and treatment of diseases
PT caused by Borrelia, particularly Lyme disease
XX
PS Claim 12; Page 201-202; 275pp; English.
XX
CC This sequence represents a Borrelia burgdorferi (Bb) protein of the
CC invention, which is suitable for use in a vaccine. The Bb polypeptides
CC can be used in vaccines for eliciting protective antibodies to members of
CC the Borrelia genus, particularly for the use against Lyme disease in
CC humans and animals. They can be used for preventing or attenuating an
CC infection caused by a member of the Borrelia genus. The products can also
CC be used for detection of members of the Borrelia genus.
XX
SQ Sequence 533 AA;

Query Match 60.4%; Score 142; DB 20; Length 533;
Best Local Similarity 68.2%; Pred. No. 7.5e-12;
Matches 30; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Qy 4 NPIDAAIGGSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
||| |||| : || |:|||||||||: |||||||:||:
Db 128 npiaaaigkgnedgaefkdemkkddqiaaaialrgmakdgkfav 171

RESULT 7
AAY20089
ID AAY20089 standard; Protein; 168 AA.
XX
AC AAY20089;
XX
DT 19-JUL-1999 (first entry)
XX
DE B. burgdorferi antigenic protein, t24-1.aa.
XX
KW Antigenic protein; vaccine; Lyme disease; infection; detection.
XX
OS Borrelia burgdorferi.
XX
PN WO9859071-A1.
XX
PD 30-DEC-1998.
XX
[illegible]18-JUN-1998; 98WO-US12718.
XX
PR 03-SEP-1997; 97US-0057483.
PR 20-JUN-1997; 97US-0050359.
PR 22-JUL-1997; 97US-0053344.
PR 22-JUL-1997; 97US-0053377.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMMUNE INC.
XX
PI Choi GH, Erwin AL, Hanson MS, Lathigra R;
XX
DR WPI; 1999-189980/16.
DR N-PSDB; AAX61786.
XX
PT New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the diagnosis, prevention and treatment of diseases
PT caused by Borrelia, particularly Lyme disease
XX
PS Claim 12; Page 193; 275pp; English.
XX
CC This sequence represents a Borrelia burgdorferi (Bb) protein of the
CC invention, which is suitable for use in a vaccine. The Bb polypeptides
CC can be used in vaccines for eliciting protective antibodies to members of
CC the Borrelia genus, particularly for the use against Lyme disease in
CC humans and animals. They can be used for preventing or attenuating an
CC infection caused by a member of the Borrelia genus. The products can also
CC be used for detection of members of the Borrelia genus.
XX
SQ Sequence 168 AA;

Query Match 56.0%; Score 131.5; DB 20; Length 168;
Best Local Similarity 66.7%; Pred. No. 5.9e-11;
Matches 30; Conservative 5; Mismatches 9; Indels 1; Gaps 1;

Qy 4 NPIDAAIG-GSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
||| |||| |: | |:|||||||||: |||||||:||:
Db 15 npiaaaigkgngdgaefdqdemkkddqiaaaialrgmakdgkfav 59

RESULT 8
AAW22676
ID AAW22676 standard; Protein; 356 AA.
XX
AC AAW22676;
XX
DT 22-FEB-1998 (first entry)
XX
DE Borrelia variable major protein (VMP)-like protein VlsE.
XX
KW Variable major protein-like sequence; VMP-like sequence; vls locus;
KW VlsE; Lyme disease; relapsing fever; therapy; diagnosis;
KW vaccine.
XX
OS Borrelia burgdorferi strain B31-5A3 (ATCC 35210).
XX
FH Key Location/Qualifiers
FT Peptide 1..19
FT /label= Sig_peptide
FT /note= "lipoprotein signal peptide"
XX
PN WO9731123-A1.
XX
PD 28-AUG-1997.
XX
PF 20-FEB-1997; 97WO-US02952.
XX
PR 21-FEB-1996; 96US-0012028.
XX
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Barbour AG, Hardham JM, Howell JK, Norris SJ, Weinstock GM;
PI Zhang J;
XX
DR WPI; 1997-435172/40.
DR N-PSDB; AAT85042.
XX
PT Nucleic acid encoding variable major protein-like peptide of
PT Borrelia - useful for recombinant production of VMP like protein or
PT peptide, or for diagnosis of Lyme disease
XX
PS Claim 2; Page 97-99; 130pp; English.
XX
CC This protein comprises the surface-exposed lipoprotein variable
CC major protein (VMP)-like protein VlsE of Borrelia burgdorferi. Its
CC sequence was deduced from an isolated vlsE gene (see AAT85042). An
CC elaborate genetic system has been identified that promotes
CC extensive antigenic variation of VlsE. An infectivity related
CC 28-kb linear plasmid, pBB28La, of B. burgdorferi B31 contains a
CC vmp-like sequence (vls) locus consisting of 15 silent vls cassettes
CC (see AAT85043) and the expressed vlsE gene. Portions of several of
CC the 15 silent vls cassette sequences, located approx. 500 bp
CC upstream of vlsE, recombine into the central vlsE cassette region
CC during infection, resulting in antigenic variation and hence immune
CC evasion, long-term survival and pathogenesis in the mammalian host.
CC Recombinant or native proteins expressed by VMP-like genes, will be
CC useful for the immunotherapy, immunoprophylaxis and immunodiagnosis

```
CC   of Lyme disease, relapsing fever and related disorders in humans
CC   and animals.  They can also be used in a Lyme disease vaccine.
XX
SQ   Sequence   356 AA;


  Query Match            55.3%;  Score 130;  DB 18;  Length 356;
  Best Local Similarity  68.9%;  Pred. No. 2.4e-10;
  Matches   31;  Conservative    4;  Mismatches    8;  Indels    2;  Gaps    2;

Qy       4 NPIDAAIGGSADRNAE-AFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
             ||| ||| |  |  ||   |:||||||||||: ||||||||:||:
Db     254 npiaaai-gdkdggaefgqdemkkddqiaaaialrgmakdgkfav 297


RESULT   9
AAB36273
ID   AAB36273 standard; peptide; 26 AA.
XX
AC   AAB36273;
XX
DT   21-FEB-2001  (first entry)
XX
DE   B. burgdorferi variable surface antigen variable domain region IR6.
XX
KW   Variable surface antigen; invariable region; VlsE; Lyme disease;
KW   Lyme borreliosis.
XX
OS   Borrelia burgdorferi.
XX
PN   WO200065064-A1.
XX
PD   02-NOV-2000.
XX
PF   25-APR-2000; 2000WO-US11085.
XX
PR   28-APR-1999;   99US-0300971.
XX
PA   (TULA ) TULANE EDUCATIONAL FUND.
XX
PI   Philipp MT,  Liang FT;
XX
DR   WPI; 2000-687350/67.
XX
PT   Novel peptides comprising an invariable 26-amino acid long region
PT   isolated from Borrelia burgdorferi (sensu lato), useful for rapid and
PT   specific diagnosis of lyme disease -
XX
PS   Claim 1; Page 59; 76pp; English.
XX
CC   The present invention describes several peptides comprised of the
CC   invariable regions IR1-IR6 of the B. burgdorferi variable surface antigen
CC   (VlsE) variable domain. These peptides can be used in the specific
CC   diagnosis, treatment and vaccination against B. burgdorferi, which causes
CC   Lyme disease (also known as Lyme borreliosis) in humans, dogs, horses,
CC   cows and other animals.
XX
SQ   Sequence   26 AA;


  Query Match            49.8%;  Score 117;  DB 21;  Length 26;
  Best Local Similarity  100.0%;  Pred. No. 7.4e-10;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      24 MKKDDQIAAAMVLRGMAKDGQFAL 47
           ||||||||||||||||||||||||
Db       1 mkkddqiaaamvlrgmakdgqfal 24


RESULT  10
AAB36280
ID   AAB36280 standard; peptide; 26 AA.
XX
AC   AAB36280;
XX
DT   21-FEB-2001  (first entry)
XX
DE   B. burgdorferi variable surface antigen modified variable domain C6.
XX
KW   Variable surface antigen; invariable region; VlsE; Lyme disease;
KW   Lyme borreliosis.
XX
OS   Borrelia burgdorferi.
OS   Synthetic.
XX
PN   WO200065064-A1.
XX
PD   02-NOV-2000.
XX
PF   25-APR-2000; 2000WO-US11085.
XX
PR   28-APR-1999;   99US-0300971.
XX
PA   (TULA ) TULANE EDUCATIONAL FUND.
XX
PI   Philipp MT,  Liang FT;
XX
DR   WPI; 2000-687350/67.
XX
PT   Novel peptides comprising an invariable 26-amino acid long region
PT   isolated from Borrelia burgdorferi (sensu lato), useful for rapid and
PT   specific diagnosis of lyme disease -
XX
PS   Example 3; Page 38; 76pp; English.
XX
CC   The present invention describes several peptides comprised of the
CC   invariable regions IR1-IR6 of the B. burgdorferi variable surface antigen
CC   (VlsE) variable domain. These peptides can be used in the specific
CC   diagnosis, treatment and vaccination against B. burgdorferi, which causes
CC   Lyme disease (also known as Lyme borreliosis) in humans, dogs, horses,
CC   cows and other animals.
XX
SQ   Sequence   26 AA;


  Query Match            49.8%;  Score 117;  DB 21;  Length 26;
  Best Local Similarity  100.0%;  Pred. No. 7.4e-10;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      24 MKKDDQIAAAMVLRGMAKDGQFAL 47
           ||||||||||||||||||||||||
Db       2 mkkddqiaaamvlrgmakdgqfal 25


RESULT  11
AAG91309
ID   AAG91309 standard; Protein; 593 AA.
XX
AC   AAG91309;
XX
DT   26-SEP-2001  (first entry)
XX
DE   C glutamicum protein fragment SEQ ID NO: 5063.
XX
KW   Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW   organic acid synthesis.
XX
OS   Corynebacterium glutamicum.
XX
PN   EP1108790-A2.
XX
PD   20-JUN-2001.
XX
PF   18-DEC-2000; 2000EP-0127688.
XX
PR   16-DEC-1999;   99JP-0377484.
```

PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
DR N-PSDB; AAH66528.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT -
XX
PS Claim 17; SEQ ID NO: 5063; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 593 AA;

Query Match 23.4%; Score 55; DB 22; Length 593;
Best Local Similarity 41.4%; Pred. No. 21;
Matches 12; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

Qy 13 SADRNAEAFDKMKKDDQIAAAMVLRGMAK 41
Db 356 sidissealdklkedakkaqasmmraakk 384

RESULT 12
AAG91164
ID AAG91164 standard; Protein; 354 AA.
XX
AC AAG91164;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum protein fragment SEQ ID NO: 4918.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
DR N-PSDB; AAH66383.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT -
XX
PS Claim 17; SEQ ID NO: 4918; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 354 AA;

Query Match 22.3%; Score 52.5; DB 22; Length 354;
Best Local Similarity 32.6%; Pred. No. 25;
Matches 14; Conservative 8; Mismatches 20; Indels 1; Gaps 1;

Qy 4 NPIDAAIGGSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFA 46
Db 74 nnielqvvgsvqggvesieslkkddidfaavpfiqlva-geia 115

RESULT 13
AAY21894
ID AAY21894 standard; Protein; 245 AA.
XX
AC AAY21894;
XX
DT 24-SEP-1999 (first entry)
XX
DE PHA acetoacetyl-CoA reductase sequence.
XX
KW Polyhydroxyalkanoate; PHA; PHA biosynthesis; PHA synthase; plastic;
KW PHA beta-ketothiolase; PHA acetoacetyl-CoA reductase; Alcaligenes latus;
KW biodegradable polymer.
XX
OS Alcaligenes latus.
XX
PN WO9936547-A1.
XX
PD 22-JUL-1999.
XX
PF 19-JAN-1999; 99WO-KR00031.
XX
PR 26-DEC-1998; 98KR-0058760.
PR 19-JAN-1998; 98KR-0001422.
PR 19-JAN-1998; 98KR-0001423.
XX
PA (GLDS) LG CHEM LTD.
XX
PI Choi J, Choo S, Han K, Hong S, Huh T, Lee SY;
PI Lee Y, Song J, Yoon H;
XX
DR WPI; 1999-444403/37.
DR N-PSDB; AAX86965, AAX86968.
XX
PT New polyhydroxyalkanoate biosynthesis-related genes used in the
PT production of biodegradable polymers

```
XX
PS   Claim 8; Page 45-56; 49pp; English.
XX
CC   The invention provides polyhydroxyalkanoate (PHA) biosynthesis-related
CC   genes. Specifically, these are the genes for PHA synthase (AAX86966),
CC   PHA beta-ketothiolase (AAX86967) and PHA acetoacetyl-CoA reductase
CC   (AAX86968), derived from  Alcaligenes latus. The genes are contained in a
CC   fragment of 6436 nucleotides (AAX86965. The genes are used to produce
CC   recombinant proteins, which are used for PHA biosynthesis. PHA can then
CC   be used to produce biodegradable polymers, which can be used as an
CC   alternative to plastic. The production cost of PHA is much higher than
CC   those of other commercially available synthetic polymers. The present
CC   invention allows the recombinant production of PHA, which results in
CC   cheaper PHA production costs. The present sequence represents the amino
CC   acid sequence of PHA acetoacetyl-CoA reductase.
XX
SQ   Sequence   245 AA;


  Query Match             21.9%;  Score 51.5;  DB 20;  Length 245;
  Best Local Similarity   34.1%;  Pred. No. 23;
  Matches   14;  Conservative    8;  Mismatches   16;  Indels    3;  Gaps    2;

Qy       8 AAIGGSA--DRNAEAFDKMKKDDQIAAAMVLR-GMAKDGQF 45
           |::|  |  |    ||:|:| :      :|   |: :|||
Db      56 asvgnvadwdstvaafekvkaehgtvdvlvnnagitrdgqf 96


RESULT  14
AAY73333
ID   AAY73333 standard; Protein; 127 AA.
XX
AC   AAY73333;
XX
DT   24-FEB-2000  (first entry)
XX
DE   HTRM clone 1760185 protein sequence.
XX
KW   HTRM; human transcriptional regulatory molecule; arteriosclerosis; AIDS;
KW   arteriosclerosis; cirrhosis; cancer; leukaemia; diabetes mellitus;
KW   Addison's disease; multiple sclerosis; rheumatoid arthritis; infection;
KW   trauma; myasthenia gravis; adenocarcinoma; immune disorder; treatment.
XX
OS   Homo sapiens.
XX
PN   WO9957144-A2.
XX
PD   11-NOV-1999.
XX
PF   04-MAY-1999;   99WO-US09935.
XX
PR   05-MAY-1998;   98US-0084254.
PR   07-AUG-1998;   98US-0095827.
PR   02-OCT-1998;   98US-0102745.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Hillman JL,  Bandman O,  Lal P,  Yue H,  Reddy R,  Tang YT;
PI   Gerstin EH,  Patterson C,  Baughn MR,  Azimzai Y,  Lu DAM;
XX
DR   WPI; 2000-052941/04.
DR   N-PSDB; AAZ52418.
XX
PT   New peptides useful for diagnosis, prevention and treatment of cancer
PT   and immune disorders  -
XX
PS   Claim 1; Page 95; 193pp; English.
XX
CC   AAY73325-Y73389 are human transcriptional regulator molecule (HTRM)
CC   protein sequences. The HTRM protein and nucleotide sequences are useful
CC   for preventing or treating disorders associated with decreased expression
CC   or activity of HTRM which include cell proliferative disorders such as
CC   arteriosclerosis and cirrhosis; cancers including adenocarcinoma and
CC   leukaemia; immune disorders such as AIDS, Addison's disease, diabetes
CC   mellitus, rheumatoid arthritis, multiple sclerosis, systemic lupus
CC   erythematosus, and myasthenia gravis; infections and trauma. Antagonists
CC   of the HTRM polypeptides are useful for treating or preventing disorders
CC   associated with  increased expression or activity of HTRMs. HTRM
CC   polypeptides, their immunogenic fragments or oligopeptides are useful
CC   for screening libraries of compounds in drug screening techniques.
CC   Polynucleotides encoding HTRM are useful for blocking the transcription
CC   of mRNA and regulating  gene function by modulating the activity of
CC   HTRM. Vectors expressing HTRM or agonists can also be used to prevent or
CC   treat disorder associated with decreased HTRM expression. Antibodies
CC   which specifically bind HTRM and polynucleotides encoding HTRM are useful
CC   for diagnosing disorders associated with the expression of HTRM,
CC   particularly in assays that detect the expression of HTRM. Nucleotide
CC   sequences encoding HTRM may be useful to generate hybridization probes
CC   useful in mapping the naturally occurring genomic sequence and to detect
CC   differences in gene sequences among normal, carrier and affected
CC   individuals. Using diagnostic assays, cancer can be detected prior to
CC   the appearance of clinical symptoms and thereby progression of cancer can
CC   be prevented by aggressive treatment or preventive measures.
XX
SQ   Sequence   127 AA;


  Query Match             21.7%;  Score 51;  DB 21;  Length 127;
  Best Local Similarity   41.2%;  Pred. No. 12;
  Matches   14;  Conservative    7;  Mismatches   13;  Indels    0;  Gaps    0;

Qy       2 LTNPIDAAIGGSADRNAEAFDKMKKDDQIAAAMV 35
           | || ||| | :| |  :| :| |:  ::|  :|
Db      82 lqnpgdaaegraakrcekaeekakeiakmaemlv 115


RESULT  15
AAW48336
ID   AAW48336 standard; Protein; 383 AA.
XX
AC   AAW48336;
XX
DT   09-JUL-1998  (first entry)
XX
DE   Brucella abortus 39 kDa antigen.
XX
KW   39 kDa antigen; detection; vaccination; infection.
XX
OS   Brucella abortus.
XX
PN   WO9808951-A1.
XX
PD   05-MAR-1998.
XX
PF   27-AUG-1997;   97WO-EP04668.
XX
PR   28-AUG-1996;   96EP-0870112.
XX
PA   (INNO-) INNOGENETICS NV.
PA   (INRG ) INST NAT RECH AGRONOMIQUE.
PA   (UYNO-) UNIV NOTRE-DAME DE LA PAIX.
XX
PI   Dubray G,  Letesson J,  Saman E,  Tibor A,  Verger J;
XX
DR   WPI; 1998-179436/16.
DR   N-PSDB; AAV17736.
XX
PT   New isolated Brucella antigens of 39 and 15 kDa - useful for
PT   developing products for detecting of, or vaccinating against
PT   brucellosis in, e.g. humans or cattle
XX
PS   Claim 2; Fig 3; 117pp; English.
XX
CC   The present sequence is a Brucella abortus 39 kDa antigen,
CC   useful for in vitro detection of, or vaccination against Brucella
CC   infection, e.g. in humans or cattle. The antigen can also be used
```

CC to detect and/or quantify the cellular immune response in an
CC individual, or detect individuals who have been in contact with
CC Brucella.
XX
SQ Sequence 383 AA;

Query Match 21.7%; Score 51; DB 19; Length 383;
Best Local Similarity 34.1%; Pred. No. 46;
Matches 15; Conservative 7; Mismatches 8; Indels 14; Gaps 2;

```
Qy       9 AIGGSADRNAEAFDKMKK----------DDQIAAAMVLRGMA 40
           |:|  :|   :|||:| |            |  :|:|||: | |
Db     185 alg--sdtmkqafdrmsklrtyvddnfsgrdwnlasamviegka 226
```

Search completed: January 10, 2002, 14:05:15
Job time: 169 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 10, 2002, 14:08:25 ; Search time 184.82 Seconds
(without alignments)
70.609 Million cell updates/sec

Title: US-09-445-803-14
Perfect score: 235
Sequence: 1 PLTNPIDAAIGGSADRNAEA.....DQIAAAMVLRGMAKDGQFAL 47

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 3148936 seqs, 277657034 residues

Total number of hits satisfying chosen parameters: 3148936

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_Main:*
1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US080_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US081_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US082_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US083_COMB.pep:*
8: /cgn2_6/ptodata/2/paa/US084_COMB.pep:*
9: /cgn2_6/ptodata/2/paa/US085_COMB.pep:*
10: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
11: /cgn2_6/ptodata/2/paa/US087_COMB.pep:*
12: /cgn2_6/ptodata/2/paa/US088_COMB.pep:*
13: /cgn2_6/ptodata/2/paa/US089_COMB.pep:*
14: /cgn2_6/ptodata/2/paa/US090_COMB.pep:*
15: /cgn2_6/ptodata/2/paa/US091_COMB.pep:*
16: /cgn2_6/ptodata/2/paa/US092_COMB.pep:*
17: /cgn2_6/ptodata/2/paa/US093_COMB.pep:*
18: /cgn2_6/ptodata/2/paa/US094_COMB.pep:*
19: /cgn2_6/ptodata/2/paa/US095_COMB.pep:*
20: /cgn2_6/ptodata/2/paa/US096_COMB.pep:*
21: /cgn2_6/ptodata/2/paa/US097_COMB.pep:*
22: /cgn2_6/ptodata/2/paa/US098_COMB.pep:*
23: /cgn2_6/ptodata/2/paa/US099_COMB.pep:*
24: /cgn2_6/ptodata/2/paa/US60_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	235	100.0	47	18	US-09-445-803-14	Sequence 14, Appl
2	176.5	75.1	349	17	US-09-300-971A-9	Sequence 9, Appli
3	176.5	75.1	349	18	US-09-445-803-2	Sequence 2, Appli
4	164.5	70.0	189	17	US-09-300-971A-10	Sequence 10, Appl
5	148.5	63.2	179	15	US-09-125-619-18	Sequence 18, Appl
6	147.5	62.8	216	15	US-09-125-619-38	Sequence 38, Appl
7	146.5	62.3	197	15	US-09-125-619-25	Sequence 25, Appl
8	145.5	61.9	216	15	US-09-125-619-46	Sequence 46, Appl
9	144.5	61.5	190	15	US-09-125-619-19	Sequence 19, Appl
10	144.5	61.5	323	17	US-09-300-971A-11	Sequence 11, Appl
11	144	61.3	189	15	US-09-125-619-26	Sequence 26, Appl
12	143.5	61.1	212	15	US-09-125-619-32	Sequence 32, Appl
13	143.5	61.1	212	15	US-09-125-619-44	Sequence 44, Appl
14	142.5	60.6	215	15	US-09-125-619-36	Sequence 36, Appl
15	142	60.4	169	15	US-09-125-619-16	Sequence 16, Appl
16	142	60.4	212	15	US-09-125-619-41	Sequence 41, Appl
17	141.5	60.2	187	15	US-09-125-619-23	Sequence 23, Appl
18	141.5	60.2	209	15	US-09-125-619-45	Sequence 45, Appl
19	140.5	59.8	190	15	US-09-125-619-27	Sequence 27, Appl
20	139.5	59.4	212	15	US-09-125-619-48	Sequence 48, Appl
21	139	59.1	194	15	US-09-125-619-28	Sequence 28, Appl
22	138	58.7	214	15	US-09-125-619-34	Sequence 34, Appl
23	136.5	58.1	169	15	US-09-125-619-24	Sequence 24, Appl
24	136.5	58.1	213	15	US-09-125-619-39	Sequence 39, Appl
25	136	57.9	195	15	US-09-125-619-17	Sequence 17, Appl
26	136	57.9	214	15	US-09-125-619-35	Sequence 35, Appl
27	134.5	57.2	208	15	US-09-125-619-47	Sequence 47, Appl
28	133.5	56.8	209	15	US-09-125-619-42	Sequence 42, Appl
29	133.5	56.8	211	15	US-09-125-619-40	Sequence 40, Appl
30	132	56.2	211	15	US-09-125-619-37	Sequence 37, Appl
31	131.5	56.0	169	15	US-09-125-619-22	Sequence 22, Appl
32	131.5	56.0	190	15	US-09-125-619-29	Sequence 29, Appl
33	131	55.7	209	15	US-09-125-619-33	Sequence 33, Appl
34	130	55.3	190	15	US-09-125-619-15	Sequence 15, Appl
35	130	55.3	212	15	US-09-125-619-31	Sequence 31, Appl
36	130	55.3	212	15	US-09-125-619-43	Sequence 43, Appl
37	130	55.3	356	1	PCT-US97-02952-2	Sequence 2, Appli
38	130	55.3	356	15	US-09-125-619-2	Sequence 2, Appli
39	130	55.3	356	15	US-09-125-619-13	Sequence 13, Appl
40	128	54.5	158	15	US-09-125-619-21	Sequence 21, Appl
41	126.5	53.8	189	15	US-09-125-619-20	Sequence 20, Appl
42	117	49.8	26	17	US-09-300-971-1	Sequence 1, Appli
43	117	49.8	26	17	US-09-300-971-8	Sequence 8, Appli
44	117	49.8	26	17	US-09-300-971A-1	Sequence 1, Appli
45	117	49.8	26	17	US-09-300-971A-8	Sequence 8, Appli

ALIGNMENTS

```
RESULT 1
US-09-445-803-14
; Sequence 14, Application US/09445803
;  GENERAL INFORMATION:
;       APPLICANT: Adminis. of Tulane Educational, Fund
;                  Philipp, Mario T.
;       TITLE OF INVENTION: Surface Antigens and Proteins Useful in
;                           Compositions for the Diagnosis and Prevention of Lyme
;                           Disease
;       NUMBER OF SEQUENCES: 14
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Howson and Howson
;            STREET: Spring House Corporate Cntr., P.O. Box 457
;            CITY: Spring House
;            STATE: Pennsylvania
;            COUNTRY: USA
;            ZIP: 19477
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/445,803
;            FILING DATE: 13-Dec-1999
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US 60/051,271
;            FILING DATE: 30-JUN-1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Bak, Mary E.
;            REGISTRATION NUMBER: 31,215
```

```
;            REFERENCE/DOCKET NUMBER: TUL2APCT
;        TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 215-540-9200
;            TELEFAX: 215-540-5818
;   INFORMATION FOR SEQ ID NO: 14:
;        SEQUENCE CHARACTERISTICS:
;            LENGTH: 47 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-09-445-803-14


  Query Match           100.0%;  Score 235;  DB 18;  Length 47;
  Best Local Similarity  100.0%;  Pred. No. 1.3e-25;
  Matches   47;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 PLTNPIDAAIGGSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
          |||||||||||||||||||||||||||||||||||||||||||||||
Db      1 PLTNPIDAAIGGSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47


RESULT   2
US-09-300-971A-9
; Sequence 9, Application US/09300971A
; GENERAL INFORMATION:
;  APPLICANT: Philipp, Mario T.
;  APPLICANT:  Liang, Fang Ting
;  TITLE OF INVENTION: Novel Peptides and Assays for the Diagnosis of Lyme Disease
;  FILE REFERENCE: TUL3USA
;  CURRENT APPLICATION NUMBER: US/09/300,971A
;  CURRENT FILING DATE:  1999-04-28
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
;   LENGTH: 349
;   TYPE: PRT
;   ORGANISM: Borrelia garinii
US-09-300-971A-9


  Query Match           75.1%;  Score 176.5;  DB 17;  Length 349;
  Best Local Similarity  82.6%;  Pred. No. 4.1e-16;
  Matches   38;  Conservative    2;  Mismatches    5;  Indels    1;  Gaps    1;

Qy      3 TNPIDAAIGGSADRN-AEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
          ||||||||||: | : | ||  ||||||||||||||||||||||||
Db    242 TNPIDAAIGGAGDNDAAAAFATMKKDDQIAAAMVLRGMAKDGQFAL 287


RESULT   3
US-09-445-803-2
; Sequence 2, Application US/09445803
;   GENERAL INFORMATION:
;        APPLICANT: Adminis. of Tulane Educational, Fund
;                   Philipp, Mario T.
;        TITLE OF INVENTION: Surface Antigens and Proteins Useful in
;                            Compositions for the Diagnosis and Prevention of Lyme
;                            Disease
;        NUMBER OF SEQUENCES: 14
;        CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Howson and Howson
;            STREET: Spring House Corporate Cntr., P.O. Box 457
;            CITY: Spring House
;            STATE: Pennsylvania
;            COUNTRY: USA
;            ZIP: 19477
;        COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/445,803
;            FILING DATE: 13-Dec-1999
;            CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US 60/051,271
;            FILING DATE: 30-JUN-1997
;        ATTORNEY/AGENT INFORMATION:
;            NAME: Bak, Mary E.
;            REGISTRATION NUMBER: 31,215
;            REFERENCE/DOCKET NUMBER: TUL2APCT
;        TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 215-540-9200
;            TELEFAX: 215-540-5818
;   INFORMATION FOR SEQ ID NO: 2:
;        SEQUENCE CHARACTERISTICS:
;            LENGTH: 349 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-445-803-2


  Query Match           75.1%;  Score 176.5;  DB 18;  Length 349;
  Best Local Similarity  82.6%;  Pred. No. 4.1e-16;
  Matches   38;  Conservative    2;  Mismatches    5;  Indels    1;  Gaps    1;

Qy      3 TNPIDAAIGGSADRN-AEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
          ||||||||||: | : | ||  ||||||||||||||||||||||||
Db    242 TNPIDAAIGGAGDNDAAAAFATMKKDDQIAAAMVLRGMAKDGQFAL 287


RESULT   4
US-09-300-971A-10
; Sequence 10, Application US/09300971A
; GENERAL INFORMATION:
;  APPLICANT: Philipp, Mario T.
;  APPLICANT:  Liang, Fang Ting
;  TITLE OF INVENTION: Novel Peptides and Assays for the Diagnosis of Lyme Disease
;  FILE REFERENCE: TUL3USA
;  CURRENT APPLICATION NUMBER: US/09/300,971A
;  CURRENT FILING DATE:  1999-04-28
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
;   LENGTH: 189
;   TYPE: PRT
;   ORGANISM: Borrelia burgdorferi
US-09-300-971A-10


  Query Match           70.0%;  Score 164.5;  DB 17;  Length 189;
  Best Local Similarity  73.9%;  Pred. No. 9.1e-15;
  Matches   34;  Conservative    7;  Mismatches    4;  Indels    1;  Gaps    1;

Qy      3 TNPIDAAIGGSADRNAEAF-DKMKKDDQIAAAMVLRGMAKDGQFAL 47
          ||||:|||| : | :| || |:|||:|||||||:||||||||||:|||
Db    128 TNPIEAAIGSTGDNDAAAFQDEMKKNDQIAAAIVLRGMAKDGEFAL 173


RESULT   5
US-09-125-619-18
; Sequence 18, Application US/09125619
; GENERAL INFORMATION:
;  APPLICANT:  NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION:  VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE:  UTSH:234
```

```
;  CURRENT APPLICATION NUMBER: US/09/125,619
;  CURRENT FILING DATE:  1999-01-27
;  NUMBER OF SEQ ID NOS:  48
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 18
;   LENGTH: 179
;   TYPE: PRT
;   ORGANISM: Borrelia burgdorferi
US-09-125-619-18


  Query Match             63.2%;  Score 148.5;  DB 15;  Length 179;
  Best Local Similarity   69.6%;  Pred. No. 1.6e-12;
  Matches   32;  Conservative    6;  Mismatches    7;  Indels    1;  Gaps    1;

Qy       3 TNPIDAAIG-GSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
           |||| |||| |: :  ||  |:||||||||||: ||||||||:||:
Db     125 TNPIAAAIGKGNEENGAEFKDEMKKDDQIAAAIALRGMAKDGKFAV 170


RESULT   6
US-09-125-619-38
; Sequence 38, Application US/09125619
; GENERAL INFORMATION:
;  APPLICANT:  NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION:  VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE:  UTSH:234
;  CURRENT APPLICATION NUMBER: US/09/125,619
;  CURRENT FILING DATE:  1999-01-27
;  NUMBER OF SEQ ID NOS:  48
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 38
;   LENGTH: 216
;   TYPE: PRT
;   ORGANISM: Borrelia burgdorferi
US-09-125-619-38


  Query Match             62.8%;  Score 147.5;  DB 15;  Length 216;
  Best Local Similarity   73.3%;  Pred. No. 2.9e-12;
  Matches   33;  Conservative    4;  Mismatches    7;  Indels    1;  Gaps    1;

Qy       4 NPIDAAIG-GSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
           ||| |||| |:||  ||  | ||||||||||: ||||||||:||:
Db     157 NPIAAAIGKGNADDGAEFGDGMKKDDQIAAAIALRGMAKDGKFAV 201


RESULT   7
US-09-125-619-25
; Sequence 25, Application US/09125619
; GENERAL INFORMATION:
;  APPLICANT:  NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION:  VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE:  UTSH:234
;  CURRENT APPLICATION NUMBER: US/09/125,619
;  CURRENT FILING DATE:  1999-01-27
;  NUMBER OF SEQ ID NOS:  48
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 25
;   LENGTH: 197
;   TYPE: PRT
;   ORGANISM: Borrelia burgdorferi
US-09-125-619-25


  Query Match             62.3%;  Score 146.5;  DB 15;  Length 197;
  Best Local Similarity   69.6%;  Pred. No. 3.5e-12;
  Matches   32;  Conservative    5;  Mismatches    8;  Indels    1;  Gaps    1;

Qy       3 TNPIDAAIG-GSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
           |||| |||| | |:  |:  | ||||||||||: ||||||||:||:
Db     143 TNPIAAAIGKGDAENGADFGDGMKKDDQIAAAIALRGMAKDGKFAV 188


RESULT   8
US-09-125-619-46
; Sequence 46, Application US/09125619
; GENERAL INFORMATION:
;  APPLICANT:  NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION:  VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE:  UTSH:234
;  CURRENT APPLICATION NUMBER: US/09/125,619
;  CURRENT FILING DATE:  1999-01-27
;  NUMBER OF SEQ ID NOS:  48
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 46
;   LENGTH: 216
;   TYPE: PRT
;   ORGANISM: Borrelia burgdorferi
US-09-125-619-46


  Query Match             61.9%;  Score 145.5;  DB 15;  Length 216;
  Best Local Similarity   69.6%;  Pred. No. 5.5e-12;
  Matches   32;  Conservative    5;  Mismatches    8;  Indels    1;  Gaps    1;

Qy       3 TNPIDAAIG-GSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
           |||| |||| |: :  ||  | ||||||||||: ||||||||:||:
Db     156 TNPIAAAIGKGNEENGAEFGDGMKKDDQIAAAIALRGMAKDGKFAV 201


RESULT   9
US-09-125-619-19
; Sequence 19, Application US/09125619
; GENERAL INFORMATION:
;  APPLICANT:  NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION:  VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE:  UTSH:234
;  CURRENT APPLICATION NUMBER: US/09/125,619
;  CURRENT FILING DATE:  1999-01-27
;  NUMBER OF SEQ ID NOS:  48
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 19
;   LENGTH: 190
;   TYPE: PRT
;   ORGANISM: Borrelia burgdorferi
US-09-125-619-19


  Query Match             61.5%;  Score 144.5;  DB 15;  Length 190;
  Best Local Similarity   71.1%;  Pred. No. 6.5e-12;
  Matches   32;  Conservative    5;  Mismatches    7;  Indels    1;  Gaps    1;

Qy       4 NPIDAAIG-GSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
           ||| |||| |:||  |:  | ||||||||||: ||||||||:||:
Db     138 NPIAAAIGKGNADDGADFGDGMKKDDQIAAAIALRGMAKDGKFAV 182
```

```
RESULT  10
US-09-300-971A-11
; Sequence 11, Application US/09300971A
; GENERAL INFORMATION:
;  APPLICANT: Philipp, Mario T.
;  APPLICANT:  Liang, Fang Ting
;  TITLE OF INVENTION: Novel Peptides and Assays for the Diagnosis of Lyme Disease
;  FILE REFERENCE: TUL3USA
;  CURRENT APPLICATION NUMBER: US/09/300,971A
;  CURRENT FILING DATE:  1999-04-28
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
;   LENGTH: 323
;   TYPE: PRT
;   ORGANISM: Borrelia burgdorferi
US-09-300-971A-11


  Query Match             61.5%;  Score 144.5;  DB 17;  Length 323;
  Best Local Similarity   71.1%;  Pred. No. 1.3e-11;
  Matches   32;  Conservative    5;  Mismatches    7;  Indels    1;  Gaps    1;

Qy       4 NPIDAAIG-GSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
           ||| |||| |:||  |:  | ||||||||||: ||||||||:||:
Db      32 NPIAAAIGKGNADDGADFGDGMKKDDQIAAAIALRGMAKDGKFAV 76


RESULT  11
US-09-125-619-26
; Sequence 26, Application US/09125619
; GENERAL INFORMATION:
;  APPLICANT:  NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION:  VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE:  UTSH:234
;  CURRENT APPLICATION NUMBER: US/09/125,619
;  CURRENT FILING DATE:  1999-01-27
;  NUMBER OF SEQ ID NOS:  48
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 26
;   LENGTH: 189
;   TYPE: PRT
;   ORGANISM: Borrelia burgdorferi
US-09-125-619-26


  Query Match             61.3%;  Score 144;  DB 15;  Length 189;
  Best Local Similarity   68.2%;  Pred. No. 7.6e-12;
  Matches   30;  Conservative    5;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       4 NPIDAAIGGSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
           ||| ||||   :  ||  |:||||||||||: ||||||||:||:
Db     136 NPIAAAIGKGNENGAEFKDEMKKDDQIAAAIALRGMAKDGKFAV 179


RESULT  12
US-09-125-619-32
; Sequence 32, Application US/09125619
; GENERAL INFORMATION:
;  APPLICANT:  NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION:  VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE:  UTSH:234
;  CURRENT APPLICATION NUMBER: US/09/125,619
;  CURRENT FILING DATE:  1999-01-27
;  NUMBER OF SEQ ID NOS:  48
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 32
;   LENGTH: 212
;   TYPE: PRT
;   ORGANISM: Borrelia burgdorferi
US-09-125-619-32


  Query Match             61.1%;  Score 143.5;  DB 15;  Length 212;
  Best Local Similarity   71.1%;  Pred. No. 1e-11;
  Matches   32;  Conservative    4;  Mismatches    8;  Indels    1;  Gaps    1;

Qy       4 NPIDAAIG-GSADRNAEAFDKMKKDDQIAAAMVI RGMAKDGQFAL 47
           ||| |||| | ||  |:  | ||||||||||: ||||||||:||:
Db     154 NPIAAAIGKGDADDGADFGDGMKKDDQIAAAIALRGMAKDGKFAV 198


RESULT  13
US-09-125-619-44
; Sequence 44, Application US/09125619
; GENERAL INFORMATION:
;  APPLICANT:  NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION:  VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE:  UTSH:234
;  CURRENT APPLICATION NUMBER: US/09/125,619
;  CURRENT FILING DATE:  1999-01-27
;  NUMBER OF SEQ ID NOS:  48
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 44
;   LENGTH: 212
;   TYPE: PRT
;   ORGANISM: Borrelia burgdorferi
US-09-125-619-44


  Query Match             61.1%;  Score 143.5;  DB 15;  Length 212;
  Best Local Similarity   71.1%;  Pred. No. 1e-11;
  Matches   32;  Conservative    4;  Mismatches    8;  Indels    1;  Gaps    1;

Qy       4 NPIDAAIG-GSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
           ||| |||| | ||  |:  | ||||||||||: ||||||||:||:
Db     154 NPIAAAIGKGDADDGADFGDGMKKDDQIAAAIALRGMAKDGKFAV 198


RESULT  14
US-09-125-619-36
; Sequence 36, Application US/09125619
; GENERAL INFORMATION:
;  APPLICANT:  NORRIS, STEVEN J.
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION:  VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE:  UTSH:234
;  CURRENT APPLICATION NUMBER: US/09/125,619
;  CURRENT FILING DATE:  1999-01-27
;  NUMBER OF SEQ ID NOS:  48
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 36
;   LENGTH: 215
;   TYPE: PRT
;   ORGANISM: Borrelia burgdorferi
```

US-09-125-619-36

Query Match 60.6%; Score 142.5; DB 15; Length 215;
Best Local Similarity 71.1%; Pred. No. 1.5e-11;
Matches 32; Conservative 4; Mismatches 8; Indels 1; Gaps 1;

```
Qy       4 NPIDAAIG-GSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
           ||| |||| | ||  ||   :||||||||||: ||||||||:||:
Db     155 NPIAAAIGKGDADDGAEFDHEMKKDDQIAAAIALRGMAKDGKFAV 199
```

RESULT 15
US-09-125-619-16
; Sequence 16, Application US/09125619
; GENERAL INFORMATION:
; APPLICANT: NORRIS, STEVEN J.
; APPLICANT: JING-REN, ZHANG
; APPLICANT: HARDHAM, JOHN M.
; APPLICANT: HOWELL, JERRILYN K.
; APPLICANT: BARBOUR, ALAN G.
; APPLICANT: WEINSTOCK, GEORGE M.
; TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
; FILE REFERENCE: UTSH:234
; CURRENT APPLICATION NUMBER: US/09/125,619
; CURRENT FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 169
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-09-125-619-16

Query Match 60.4%; Score 142; DB 15; Length 169;
Best Local Similarity 68.2%; Pred. No. 1.3e-11;
Matches 30; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

```
Qy       4 NPIDAAIGGSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
           ||| ||||   :  ||  |:||||||||||: ||||||||:||:
Db     118 NPIAAAIGKGNEDGAEFKDEMKKDDQIAAAIALRGMAKDGKFAV 161
```

Search completed: January 10, 2002, 14:08:25
Job time: 359 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 10, 2002, 14:04:01 ; Search time 38 Seconds
(without alignments)
1343.396 Million cell updates/sec

Title: US-09-445-803-2
Perfect score: 1708
Sequence: 1 KNNDHDNHKGTVKNAVDMAK.....GNGATAKGGDAKSVNGIAKG 349

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_17:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	465	27.2	192	2	Q9RHW5	Q9rhw5 borrelia bu
2	459	26.9	187	2	Q9RHW3	Q9rhw3 borrelia bu
3	454	26.6	223	2	Q9RHW7	Q9rhw7 borrelia bu
4	452.5	26.5	189	2	Q9RHW6	Q9rhw6 borrelia bu
5	452.5	26.5	190	2	Q9RHW4	Q9rhw4 borrelia bu
6	438	25.6	224	2	Q9RF40	Q9rf40 borrelia bu
7	437.5	25.6	356	2	O06878	O06878 borrelia bu
8	434.5	25.4	217	2	O68375	O68375 borrelia bu
9	433.5	25.4	358	2	O07055	O07055 borrelia bu
10	429	25.1	212	2	O68361	O68361 borrelia bu
11	428	25.1	212	2	O68368	O68368 borrelia bu
12	426	24.9	222	2	Q9RF44	Q9rf44 borrelia bu
13	426	24.9	223	2	Q9RF43	Q9rf43 borrelia bu
14	423.5	24.8	209	2	O07056	O07056 borrelia bu
15	422	24.7	214	2	O68358	O68358 borrelia bu
16	421.5	24.7	217	2	O68376	O68376 borrelia bu
17	421.5	24.7	356	2	O07053	O07053 borrelia bu
18	420.5	24.6	213	2	O68373	O68373 borrelia bu
19	420	24.6	210	2	O07054	O07054 borrelia bu
20	419	24.5	210	2	O68369	O68369 borrelia bu
21	419	24.5	218	2	Q9RF39	Q9rf39 borrelia bu
22	418.5	24.5	213	2	O68354	O68354 borrelia bu
23	418	24.5	212	2	O68360	O68360 borrelia bu
24	418	24.5	214	2	O68340	O68340 borrelia bu
25	417.5	24.4	211	2	O68366	O68366 borrelia bu
26	417.5	24.4	221	2	Q9RF38	Q9rf38 borrelia bu
27	417	24.4	212	2	O07062	O07062 borrelia bu
28	417	24.4	214	2	O07058	O07058 borrelia bu
29	417	24.4	216	2	O68355	O68355 borrelia bu
30	415	24.3	214	2	O68303	O68303 borrelia bu
31	414	24.2	216	2	O68349	O68349 borrelia bu
32	413.5	24.2	209	2	O07065	O07065 borrelia bu
33	413.5	24.2	215	2	O68356	O68356 borrelia bu
34	413.5	24.2	217	2	O07061	O07061 borrelia bu
35	413	24.2	214	2	O68362	O68362 borrelia bu
36	413	24.2	216	2	O68365	O68365 borrelia bu
37	412.5	24.2	211	2	O68367	O68367 borrelia bu
38	412.5	24.2	211	2	O68379	O68379 borrelia bu
39	412.5	24.2	213	2	O68348	O68348 borrelia bu
40	412.5	24.2	213	2	O68370	O68370 borrelia bu
41	412	24.1	214	2	O68357	O68357 borrelia bu
42	411.5	24.1	215	2	O68364	O68364 borrelia bu
43	411	24.1	222	2	Q9RF45	Q9rf45 borrelia bu
44	410.5	24.0	211	2	O68351	O68351 borrelia bu
45	410.5	24.0	221	2	Q9R349	Q9r349 borrelia bu

ALIGNMENTS

RESULT 1
Q9RHW5
ID Q9RHW5 PRELIMINARY; PRT; 192 AA.
AC Q9RHW5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE VLSF8 PROTEIN (FRAGMENT).
GN VLSF8.
OS Borrelia burgdorferi (Lyme disease spirochete).
OG Plasmid 20-kb borrelial plasmid.
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=297;
RX MEDLINE=98181050; PubMed=9514637;
RA Kawabata H., Myouga F., Inagaki Y., Murai N., Watanabe H.;
RT "Genetic and immunological analyses of Vls (VMP-like sequences) of
RT Borrelia burgdorferi.";
RL Microb. Pathog. 24:155-166(1998).
DR EMBL; AB011065; BAA87888.1; -.
DR InterPro; IPR000680; Borrelia_lipo.
DR Pfam; PF00921; Lipoprotein_2; 1.
KW Plasmid.
FT NON_TER 1 1
FT NON_TER 192 192
SQ SEQUENCE 192 AA; 18638 MW; F45DE1DF40A81A12 CRC64;

Query Match 27.2%; Score 465; DB 2; Length 192;
Best Local Similarity 65.2%; Pred. No. 4.3e-15;
Matches 118; Conservative 13; Mismatches 38; Indels 12; Gaps 8;

```
Qy    117 AAVAASAATGNAAIGDVVN-GDVAKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKL-- 173
          || |||  ||:  || ||     | ||||:  |||||| |||||| |||||  :||||
Db      3 AATAASGGTGD-MIGKVVKVTGAAAAKGGEEKSVNGIASGIKGIVAAAEKA-GEEGKLES 60

Qy    174 NAAGAEGTTNADAGKLFV---KNAGNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDG 230
          |||| |  | ||||||     || | || |||| |||||:||||:||||||| ||  |
Db     61 EAAGAAGEANEDAGKLFAAKNANAANGGGGAGDAEKAAAAVSAVSGKQILKAIVDAA--G 118
```

```
Qy    231 GEKQG-KKAADRTNPIDAAIGGAGDNDAAA-AFATMKKDDQIAAAMVLRGMAKDGQFALK 288
          |::|     : ||||:|||| |||||||  :   |||:|||||||:|||||||||:|||| 
Db    119 KEEKGVADVKEATNPIEAAIGSTGDNDAAAFSQDEMKKNDQIAAAIVLRGMAKDGEFALK 178

Qy    289 D 289
          |
Db    179 D 179


RESULT   2
Q9RHW3
ID   Q9RHW3      PRELIMINARY;      PRT;   187 AA.
AC   Q9RHW3;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   VLSF16 PROTEIN (FRAGMENT).
GN   VLSF16.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OG   Plasmid 20-kb borrelial plasmid.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=297;
RX   MEDLINE=98181050; PubMed=9514637;
RA   Kawabata H., Myouga F., Inagaki Y., Murai N., Watanabe H.;
RT   "Genetic and immunological analyses of Vls (VMP-like sequences) of
RT   Borrelia burgdorferi.";
RL   Microb. Pathog. 24:155-166(1998).
DR   EMBL; AB011067; BAA87890.1; -.
DR   InterPro; IPR000680; Borrelia_lipo.
DR   Pfam; PF00921; Lipoprotein_2; 1.
KW   Plasmid.
FT   NON_TER       1      1
FT   NON_TER     187    187
SQ   SEQUENCE   187 AA;  18451 MW;  5E81030A80396C96 CRC64;


  Query Match             26.9%;  Score 459;  DB 2;  Length 187;
  Best Local Similarity   65.9%;  Pred. No. 8e-15;
  Matches  118;  Conservative   13;  Mismatches   38;  Indels   10;  Gaps    8;

Qy    120 AASAATGNA-AIGDVVNGDV-AKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAG 177
          ||:||||:   || ||  |    |||||| |||||| ||||||:|||||  :|||| :
Db      3 AATAATGDKDMIGKVVKADTDTAAKGGDATSVNGIASGIKGIVEAAEKA-GEEGKLESEA 61

Qy    178 AEGTTNADAGKLFV-KNAGNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQG- 235
          | |  |||||||  ||  | || | || |||||:||||:|||||| ||  | |::|
Db     62 A-GDGNADAGKLFAKKNDDNGGGGAEDAEKAAAAVSAVSGKQILKAIVDAA--GKEEKGV 118

Qy    236 KKAADRTNPIDAAIGGAGDNDAAAAF--ATMKKDDQIAAAMVLRGMAKDGQFALKDAAA 292
          :   | |||| ||||   |  |||||  | |||:|||||||:||| |||||:|||||| |
Db    119 EDVKDATNPIAAAIGSTDDQKAAAAFSQAGMKKNDQIAAAIVLRKMAKDGEFALKDADA 177


RESULT   3
Q9RHW7
ID   Q9RHW7      PRELIMINARY;      PRT;   223 AA.
AC   Q9RHW7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   VLSII PROTEIN (FRAGMENT).
GN   VLSII.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OG   Plasmid 20-kb borrelial plasmid.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=297;
RX   MEDLINE=98181050; PubMed=9514637;
RA   Kawabata H., Myouga F., Inagaki Y., Murai N., Watanabe H.;
RT   "Genetic and immunological analyses of Vls (VMP-like sequences) of
RT   Borrelia burgdorferi.";
RL   Microb. Pathog. 24:155-166(1998).
DR   EMBL; AB011063; BAA87886.1; -.
DR   InterPro; IPR000680; Borrelia_lipo.
DR   Pfam; PF00921; Lipoprotein_2; 1.
KW   Plasmid.
FT   NON_TER     223    223
SQ   SEQUENCE   223 AA;  22053 MW;  47080F3C3C45C82A CRC64;


  Query Match             26.6%;  Score 454;  DB 2;  Length 223;
  Best Local Similarity   58.4%;  Pred. No. 1.6e-14;
  Matches  122;  Conservative   20;  Mismatches   43;  Indels   24;  Gaps   10;

Qy    116 AAAVAASAATGN-AAIGDVV---NGDVAKAKGGDAASVNGIAKGIKGIVDAAEKADAKEG 171
          ||  ||:||:|:   || ||   |   | ||||:  |||||| ||||||:|||||  |||
Db      4 AATEAATAASGDKEMIGKVVKVTNAGAAAAKGGEEKSVNGIASGIKGIVEAAEKA-GKEG 62

Qy    172 KLNA--AGAEGTTNAD------AGKLFVK--NAGNVGGEAGDAGKAAAAVAAVSGEQILK 221
          || :  | | |  | |       |||||  |  || | || |  | ||||||:||||:||||
Db     63 KLESEEAAAAGDGNKDACGNENAGKLFAKNANAAANGGGGAAAAEKAAAAVSAVSGKQILK 122

Qy    222 AIVHAAKDGGEKQG-KKAADRTNPIDAAIGGAGDNDAAA-AFATMKKDDQIAAAMVLRGM 279
          ||| ||  | |::|     : ||||:||||  |||||||  :   |||:|||||:|||||
Db    123 AIVDAA--GKEEKGVADVKEATNPIEAAIGSTGDNDAAAFSQDEMKKNDQIAAAIVLRGM 180

Qy    280 AKDGQFALKD-----AAAAHEGTVKNAVD 303
          ||||:|||||     |    : ||::||:
Db    181 AKDGEFALKDNEHDKAEKGLKSTVESAVN 209


RESULT   4
Q9RHW6
ID   Q9RHW6      PRELIMINARY;      PRT;   189 AA.
AC   Q9RHW6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   VLSF2 PROTEIN (FRAGMENT).
GN   VLSF2.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OG   Plasmid 20-kb borrelial plasmid.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=297;
RX   MEDLINE=98181050; PubMed=9514637;
RA   Kawabata H., Myouga F., Inagaki Y., Murai N., Watanabe H.;
RT   "Genetic and immunological analyses of Vls (VMP-like sequences) of
RT   Borrelia burgdorferi.";
RL   Microb. Pathog. 24:155-166(1998).
DR   EMBL; AB011064; BAA87887.1; -.
DR   InterPro; IPR000680; Borrelia_lipo.
DR   Pfam; PF00921; Lipoprotein_2; 1.
KW   Plasmid.
FT   NON_TER       1      1
FT   NON_TER     189    189
SQ   SEQUENCE   189 AA;  18276 MW;  7487CB3B9FF4E4B5 CRC64;


  Query Match             26.5%;  Score 452.5;  DB 2;  Length 189;
  Best Local Similarity   63.9%;  Pred. No. 1.6e-14;
  Matches  117;  Conservative   14;  Mismatches   39;  Indels   13;  Gaps    8;

Qy    117 AAVAASAATGNAAIGDVVN-GDVAKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKL-- 173
          || ||:  ||:  || ||     | ||||:  |||||| |||||| |||||  :||||
Db      3 AATAATGGTGD-MIGKVVKVTGAAAAKGGEEKSVNGIASGIKGIVAAAEKA-GEEGKLES 60

Qy    174 NAAGAEGTTNADAGKLFV---KNAGNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAA-KD 229
          |||| |  | || |||     || | || | || ||||||:||||:||||||| || |:
```

```
Db     61 EAAGAAGEANEDAAKLFAAKNANAANGGGGAPDAEKAAAAVSAVSGKQILKAIVDAAGKE 120

Qy    230 GGEKQGKKAADRTNPIDAAIGGAGDNDAAAAFATMKKDDQIAAAMVLRGMAKDGQFALKD 289
          ||:| || : |||| ||||  |:   ||||: |||:||||||:||| |||||:|||||
Db    121 --EKKGAKAEEATNPIAAAIGSTGEK--AAAFSHMKKNDQIAAAIVLRKMAKDGEFALKD 176

Qy    290 AAA 292
          | |
Db    177 ADA 179


RESULT   5
Q9RHW4
ID   Q9RHW4      PRELIMINARY;      PRT;   190 AA.
AC   Q9RHW4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   VLSF15 PROTEIN (FRAGMENT).
GN   VLSF15.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OC   Plasmid 20-kb borrelial plasmid.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=297;
RX   MEDLINE=98181050; PubMed=9514637;
RA   Kawabata H., Myouga F., Inagaki Y., Murai N., Watanabe H.;
RT   "Genetic and immunological analyses of Vls (VMP-like sequences) of
RT   Borrelia burgdorferi.";
RL   Microb. Pathog. 24:155-166(1998).
DR   EMBL; AB011066; BAA87889.1; -.
DR   InterPro; IPR000680; Borrelia_lipo.
DR   Pfam; PF00921; Lipoprotein_2; 1.
KW   Plasmid.
FT   NON_TER       1      1
FT   NON_TER     190    190
SQ   SEQUENCE   190 AA;  18748 MW;  DF5EA566DA6D29CB CRC64;


  Query Match             26.5%;  Score 452.5;  DB 2;  Length 190;
  Best Local Similarity   62.2%;  Pred. No. 1.6e-14;
  Matches  112;  Conservative   18;  Mismatches   41;  Indels    9;  Gaps    7;

Qy    120 AASAATGNA--AIGDV--VNGDVAKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNA 175
          ||:||:|     || |  ||   | ||||:  |||||| |||||| |||||  :|||| :
Db      3 AATAASGGKEDMIGKVVKVNAAGAAAKGGEETSVNGIASGIKGIVTAAEKA-GEEGKLKS 61

Qy    176 -AGAEGTTNADAGKLFVKNAGNVGGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQ 234
           |  :|  | |||||| |   : ||:| || |||||||:||||:||||||| |||  |||:
Db     62 EAAGDGEANEDAGKLFAKK-NDTGGDAKDAEKAAAAVSAVSGKQILKAIVDAAKGDGEKK 120

Qy    235 G-KKAADRTNPIDAAIGGAGDNDAAA-AFATMKKDDQIAAAMVLRGMAKDGQFALKDAAA 292
          |     :  ||| |||| |||||||  :   |||:||:|||:|||||||||:||||:| |
Db    121 GVDDVKEAKNPIAAAIGSTGDNDAAAFSKDEMKKNDQVAAAIVLRGMAKDGEFALKNADA 180


RESULT   6
Q9RF40
ID   Q9RF40      PRELIMINARY;      PRT;   224 AA.
AC   Q9RF40;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   VLSE (FRAGMENT).
GN   VLSE.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W12M;
RA   Iyer R., Hardham J.M., Wormser G.P., Schwartz I., Norris S.J.;
RT   "Conservation and heterogeneity of vlsE among human and tick isolates
RT   of Borrelia burgdorferi.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF201679; AAF25655.1; -.
DR   InterPro; IPR000680; Borrelia_lipo.
DR   Pfam; PF00921; Lipoprotein_2; 1.
FT   NON_TER       1      1
FT   NON_TER     224    224
SQ   SEQUENCE   224 AA;  21807 MW;  FB0900B833615E7B CRC64;


  Query Match             25.6%;  Score 438;  DB 2;  Length 224;
  Best Local Similarity   50.4%;  Pred. No. 9e-14;
  Matches  120;  Conservative   26;  Mismatches   68;  Indels   24;  Gaps    9;

Qy     79 GKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGDVVNGDV 138
          || |:||: ||         |: |  :  :  |    |   |  |::| ||||:||  |
Db      4 GKPDSTGSVGTA--------VEGAIKEVSELLDKLVKAVKTAEGASSGTAAIGEVV-ADA 54

Qy    139 AKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNA-AGAEGTTNADAGKLFVKNAGNV 197
            ||  | ||| |||||| ||:||  ::    || | | |:| :| :|||| |   |
Db     55 GAAKVADKASVTGIAKGIKEIVEAAGGSE----KLKAVAAAKGESNKEAGKLFGKAGANN 110

Qy    198 GGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDG-GEKQGKKAADRTNPIDAAIGGAGDND 256
           |::  | ||| ||:|||||||| ||| ||  |  |::||| ||  ||| ||| | || :
Db    111 AGDSEAASKAAGAVSAVSGEQILSAIVTAAGAGEAEQEGKKPADAKNPIAAAI-GKGDEE 169

Qy    257 AAAAFA--TMKKDDQIAAAMVLRGMAKDGQFALKDAAAAHEGTVKNAVDIIKAAAEAA 312
            | |    ||||||||||: ||||||||:||:||     :|  : |   || |||:|
Db    170 DGAEFGKDEMKKDDQIAAAIALRGMAKDGKFAVKD---DEKGRAEGA---IKGAAESA 221


RESULT   7
O06878
ID   O06878      PRELIMINARY;      PRT;   356 AA.
AC   O06878;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   VMP-LIKE VLSE.
GN   VLSE.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B31 (ATCC 35210);
RX   MEDLINE=97262068; PubMed=9108482;
RA   Zhang J.R., Hardham J.M., Barbour A.G., Norris S.J.;
RT   "Antigenic variation in Lyme disease borreliae by promiscuous
RT   recombination of VMP-like sequence cassettes.";
RL   Cell 89:275-285(1997).
DR   EMBL; U76405; AAC45733.1; -.
DR   InterPro; IPR000680; Borrelia_lipo.
DR   Pfam; PF00921; Lipoprotein_2; 1.
SQ   SEQUENCE   356 AA;  36012 MW;  304035731B2AA2BE CRC64;


  Query Match             25.6%;  Score 437.5;  DB 2;  Length 356;
  Best Local Similarity   46.9%;  Pred. No. 1.5e-13;
  Matches  127;  Conservative   26;  Mismatches   87;  Indels   31;  Gaps   11;

Qy     79 GKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGDVVNGDV 138
          || |:||: ||         |: |  :  :  |    |   |  |::| ||||:||  |
Db    103 GKPDSTGSVGTA--------VEGAIKEVSELLDKLVKAVKTAEGASSGTAAIGEVV-ADA 153

Qy    139 AKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNA-AGAEGTTNADAGKLFVKNAGNV 197
            ||  | ||| ||||||| ||:||  ::    || | | |:|  |  |||| |
Db    154 DAAKVADKASVKGIAKGIKEIVEAAGGSE----KLKAVAAAKGENNKGAGKLFGKAGAAA 209

Qy    198 GGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAADRTNPIDAAIGGAGDNDA 257
           |::  | ||| ||:|||||||| ||| || |  |: |||  :  ||| |||   || |
```

```
Db    210 HGDSEAASKAAGAVSAVSGEQILSAIVTAA-DAAEQDGKKPEEAKNPIAAAI---GDKDG 265

Qy    258 AAAFA--TMKKDDQIAAAMVLRGMAKDGQFALKDAAAAHEGTVKNAVDIIKAAAEAA--S 313
          | |    ||||||||||: ||||||||:||:||      |  : |   || |||:|
Db    266 GAEFGQDEMKKDDQIAAAIALRGMAKDGKFAVKD------GEKEKAEGAIKGAAESAVRK 319

Qy    314 AASAATGSAAIGDVVNGNGATAKGGDAKSVN 344
            | ||   ||| |: :|    |   |: :
Db    320 VLGAITG--LIGDAVS-SGLRKVGDSVKAAS 347


RESULT   8
O68375
ID   O68375      PRELIMINARY;      PRT;   217 AA.
AC   O68375;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   VMP-LIKE SEQUENCE PROTEIN VSLE (FRAGMENT).
GN   VLSE.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OG   Plasmid lp28-1.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B31-1500B;
RA   Zhang J.-R., Norris S.J.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF034523; AAC31352.1; -.
DR   InterPro; IPR000680; Borrelia_lipo.
DR   Pfam; PF00921; Lipoprotein_2; 1.
KW   Plasmid.
FT   NON_TER       1      1
FT   NON_TER     217    217
SQ   SEQUENCE   217 AA;  21109 MW;  7B34DC72AA30D83D CRC64;

  Query Match             25.4%;  Score 434.5;  DB 2;  Length 217;
  Best Local Similarity   49.6%;  Pred. No. 1.3e-13;
  Matches  115;  Conservative   22;  Mismatches   68;  Indels   27;  Gaps    7;

Qy     79 GKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGDVVNGDV 138
          || |:||: ||         |: |  :  :  |    |   |  |::|  ||| ||: |
Db      4 GKPDSTGSVGTA--------VEGAIKEVSELLDKLVKAVKTAEGASSGTDAIGKVVDNDA 55

Qy    139 AKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNA-AGAEGTTNADAGKLFVKNAGNV 197
            ||  |  || ||||||| ||:|| ::    || | | | |  |  |||| |
Db     56 GAAKAADKESVKGIAKGIKEIVEAAGGSE----KLKAVAAATGENNKKAGKLFGK----- 106

Qy    198 GGEAGDAG------KAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAADRTNPIDAAIGG 251
           |:||||       ||| ||:|||||||| ||| || | |:::||| || |||| ||||
Db    107 AGDAGDAGDSEAASKAAGAVSAVSGEQILSAIVKAA-DAGDQEGKKPADATNPIAAAIGK 165

Qy    252 AGDNDAAAAFATMKKDDQIAAAMVLRGMAKDGQFALKDA--AAAHEGTVKNA 301
            :: |      |||||||||: ||||||||:||:||       || :| |
Db    166 GNEDGAEFNQDGMKKDDQIAAAIALRGMAKDGKFAVKDGDEKGKAEGAIKGA 217


RESULT   9
O07055
ID   O07055      PRELIMINARY;      PRT;   358 AA.
AC   O07055;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   VMP-LIKE SEQUENCE PROTEIN VLSE.
GN   VLSE.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OG   Plasmid lp28-1.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B31-M1E4C;
RX   MEDLINE=97262068; PubMed=9108482;
RA   Zhang J.R., Hardham J.M., Barbour A.G., Norris S.J.;
RT   "Antigenic variation in Lyme disease borreliae by promiscuous
RT   recombination of VMP-like sequence cassettes.";
RL   Cell 89:275-285(1997).
DR   EMBL; U84556; AAC45764.1; -.
DR   InterPro; IPR000680; Borrelia_lipo.
DR   Pfam; PF00921; Lipoprotein_2; 1.
KW   Plasmid.
SQ   SEQUENCE   358 AA;  36009 MW;  6D22AE9AAA076EEC CRC64;


  Query Match             25.4%;  Score 433.5;  DB 2;  Length 358;
  Best Local Similarity   46.4%;  Pred. No. 2.4e-13;
  Matches  128;  Conservative   26;  Mismatches   83;  Indels   39;  Gaps   12;

Qy     79 GKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGDVVNGDV 138
          || |:||: ||         |: |  :  :  |    |   |  |::| ||||:|| |
Db    103 GKPDSTGSVGTA--------VEGAIKEVSELLDKLVKAVKTAEGASSGTAAIGEVV-ADA 153

Qy    139 AKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGAEGTTNADAGKLFVKNAGNVG 198
          |||   |  || ||||||| ||:||  ::    ||  | | | :|  ||||    | |
Db    154 AKA--ADKDSVKGIAKGIKEIVEAAGGSE----KLKVAAATGESNKGAGKLF----GKVD 203

Qy    199 GEAGD---AGKAAAAVAAVSGEQILKAIV---HAAKDGGEKQGKKAADRTNPIDAAIGGA 252
          | |||   | ||| ||:|||||||| |||    ||    :::||| || |||| ||||
Db    204 GAAGDSEAASKAAGAVSAVSGEQILSAIVTAAGAAASEADQEGKKPADATNPIAAAIGKG 263

Qy    253 GDNDAAAAFATMKKDDQIAAAMVLRGMAKDGQFALK--DAAAAHEGTVKNAVDIIKAAAE 310
           : : |      |||||||||: |||||||||:||:|  |     ||         || |||
Db    264 NEENGAEFGDGMKKDDQIAAAIALRGMAKDGKFAVKNDDEKGKAEGA-------IKGAAE 316

Qy    311 AA--SAASAATGSAAIGDVVNGNGATAKGGDAKSVN 344
          :|      | ||   ||| |: :|    |   |: :
Db    317 SAVRKVLGAITG--LIGDAVS-SGLRKVGDSVKAAS 349


RESULT  10
O68361
ID   O68361      PRELIMINARY;      PRT;   212 AA.
AC   O68361;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   VMP-LIKE SEQUENCE PROTEIN VSLE (FRAGMENT).
GN   VLSE.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OG   Plasmid lp28-1.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B31-1394D;
RA   Zhang J.-R., Norris S.J.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF034509; AAC31338.1; -.
DR   InterPro; IPR000680; Borrelia_lipo.
DR   Pfam; PF00921; Lipoprotein_2; 1.
KW   Plasmid.
FT   NON_TER       1      1
FT   NON_TER     212    212
SQ   SEQUENCE   212 AA;  20488 MW;  A98950CCEF789BDF CRC64;


  Query Match             25.1%;  Score 429;  DB 2;  Length 212;
  Best Local Similarity   50.0%;  Pred. No. 2.2e-13;
  Matches  113;  Conservative   24;  Mismatches   69;  Indels   20;  Gaps    7;

Qy     79 GKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGDVVNGDV 138
          || |:||: ||         |: |  :  :  |    |   |  |::| ||||:||: |
Db      4 GKPDSTGSVGTA--------VEGAIKEVSELLDKLVKAVKTAEGASSGTAAIGEVVDND- 54
```

```
Qy    139 AKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGAEGTTNADAGKLFVKNAGNVG 198
          ||  | ||| ||||||| ||:||  ::    ||  | | |  |  ||||| |   :
Db     55 --AKVADKASVTGIAKGIKEIVEAAGGSE----KLKVAAATGENNKGAGKLFGKAGADAH 108

Qy    199 GEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAADRTNPIDAAIGGAGDNDAA 258
          |::  | ||| ||:|||||||| ||| || |  |: |:| |:  ||| ||| | || ||
Db    109 GDSEAASKAAGAVSAVSGEQILSAIVKAA-DAAEQDGEKPAEAKNPIAAAI-GKGDGDAG 166

Qy    259 AAFAT-MKKDDQIAAAMVLRGMAKDGQFALK--DAAAAHEGTVKNA 301
          | |   ||||||||||: ||||||||:||:|  :     || :| |
Db    167 ADFGDGMKKDDQIAAAIALRGMAKDGKFAVKGNNEKGKAEGAIKGA 212


RESULT  11
O68368
ID   O68368      PRELIMINARY;      PRT;   212 AA.
AC   O68368;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   VMP-LIKE SEQUENCE PROTEIN VSLE (FRAGMENT).
GN   VLSE.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OG   Plasmid lp28-1.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B31-1395F;
RA   Zhang J.-R., Norris S.J.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF034516; AAC31345.1; -.
DR   InterPro; IPR000680; Borrelia_lipo.
DR   Pfam; PF00921; Lipoprotein_2; 1.
KW   Plasmid.
FT   NON_TER       1      1
FT   NON_TER     212    212
SQ   SEQUENCE   212 AA;  20645 MW;  C7DAE7377957B175 CRC64;


  Query Match             25.1%;  Score 428;  DB 2;  Length 212;
  Best Local Similarity   50.4%;  Pred. No. 2.4e-13;
  Matches  114;  Conservative   20;  Mismatches   72;  Indels   20;  Gaps    7;

Qy     79 GKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGDVVNGDV 138
          || |:||: ||          |: |  :  :  |    |   |  |::| ||||:||  |
Db      4 GKPDSTGSVGTA--------VEGAIKEVSELLDKLVKAVKTAEGASSGTAAIGEVV-ADA 54

Qy    139 AKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGAEGTTNADAGKLFVKNAGNVG 198
          ||  | ||| ||||||| ||:||  ::    || || |||  |  ||||| |
Db     55 DAAKVADKASVKGIAKGIKEIVEAAGGSE----KLKAAAAEGENNKGAGKLFGKAGAAAH 110

Qy    199 GEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAADRTNPIDAAIGGAGDNDAA 258
          |::  | ||| ||:|||||||| ||| || |  |: |||  :  ||| |||   || |
Db    111 GDSEAASKAAGAVSAVSGEQILSAIVTAA-DAAEQDGKKPEEAKNPIAAAI---GDKDGG 166

Qy    259 AAFA--TMKKDDQIAAAMVLRGMAKDGQFALKDAAAAH-EGTVKNA 301
          | |    ||||||||||: ||||||||:||:||      || :| |
Db    167 AEFGQDEMKKDDQIAAAIALRGMAKDGKFAVKDGEKEKAEGAIKGA 212


RESULT  12
Q9RF44
ID   Q9RF44      PRELIMINARY;      PRT;   222 AA.
AC   Q9RF44;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   VLSE (FRAGMENT).
GN   VLSE.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OG   Plasmid 32 kb.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B294;
RA   Iyer R., Hardham J.M., Wormser G.P., Schwartz I., Norris S.J.;
RT   "Conservation and heterogeneity of vlsE among human and tick isolates
RT   of Borrelia burgdorferi.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF201675; AAF25651.1; -.
DR   InterPro; IPR000680; Borrelia_lipo.
DR   Pfam; PF00921; Lipoprotein_2; 1.
KW   Plasmid.
FT   NON_TER       1      1
FT   NON_TER     222    222
SQ   SEQUENCE   222 AA;  21635 MW;  6E4782BB9A092ACF CRC64;


  Query Match             24.9%;  Score 426;  DB 2;  Length 222;
  Best Local Similarity   50.0%;  Pred. No. 3.2e-13;
  Matches  115;  Conservative   21;  Mismatches   76;  Indels   18;  Gaps    7;

Qy     78 EGKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGDVVNGD 137
          :|| |:||: ||          |: |  :  :  |    |   |  | :|  |||:||  |
Db      2 DGKPDSTGSVGTA--------VEGAIKEVSELLDKLVKAVKTAEGAPSGTDAIGEVV-AD 52

Qy    138 VAKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGAEGTTNADAGKLFVKNAGNV 197
             ||  | ||| ||||||| ||:||  ::    || || | |  |  ||||| |
Db     53 AGAAKAADKASVTGIAKGIKEIVEAAGGSE----KLKAAAAAGENNKKAGKLFGKVDAAH 108

Qy    198 GGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDG-GEKQGKKAADRTNPIDAAIGGAGDND 256
             |::  | ||| ||:|||||||| ||| ||  |  |: |:| ||  ||| ||| | || |
Db    109 AGDSEAASKAAGAVSAVSGEQILSAIVKAAAAAGAAEQDGEKPADAKNPIAAAI-GKGDAD 167

Qy    257 AAAAF-ATMKKDDQIAAAMVLRGMAKDGQFALK--DAAAAHEGTVKNAVD 303
          | | |   ||||||||||: ||||||||:||:|  |     || :| | :
Db    168 AGADFNHEMKKDDQIAAAIALRGMAKDGKFAVKNDDEKGKAEGAIKGAAE 217


RESULT  13
Q9RF43
ID   Q9RF43      PRELIMINARY;      PRT;   223 AA.
AC   Q9RF43;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   VLSE (FRAGMENT).
GN   VLSE.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OG   Plasmid 32 kb.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B294;
RA   Iyer R., Hardham J.M., Wormser G.P., Schwartz I., Norris S.J.;
RT   "Conservation and heterogeneity of vlsE among human and tick isolates
RT   of Borrelia burgdorferi.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF201676; AAF25652.1; -.
DR   InterPro; IPR000680; Borrelia_lipo.
DR   Pfam; PF00921; Lipoprotein_2; 1.
KW   Plasmid.
FT   NON_TER       1      1
FT   NON_TER     223    223
SQ   SEQUENCE   223 AA;  21636 MW;  0B344689C00DC3A0 CRC64;


  Query Match             24.9%;  Score 426;  DB 2;  Length 223;
  Best Local Similarity   50.2%;  Pred. No. 3.2e-13;
  Matches  115;  Conservative   21;  Mismatches   75;  Indels   18;  Gaps    7;

Qy     79 GKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGDVVNGDV 138
```

```
          || |:||: ||         |: |  :  :  |    |   |  |::|  |||:||  |
Db      4 GKPDSTGSVGTA--------VEGAIKEVSELLDKLVKAVKTAEGASSGTDAIGEVV-ADA 54

Qy    139 AKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGAEGTTNADAGKLFVKNAGNVG 198
            ||  | ||| |||||||| ||:||  ::    || || | |  |  |||| |
Db     55 GAAKAADKASVTGIAKGIKEIVEAAGGSE----KLKAAAAAGENNKKAGKLFGKVDAAHA 110

Qy    199 GEAGDAGKAAAAVAAVSGEQILKAIVHAAKDG-GEKQGKKAADRTNPIDAAIGGAGDNDA 257
          |::  | ||| ||:||||||||| ||| ||  |  |: |:| ||  ||| ||| | || ||
Db    111 GDSEAASKAAGAVSAVSGEQILSAIVKAAAAGAAEQDGEKPADAKNPIAAAI-GKGDADA 169

Qy    258 AAAF-ATMKKDDQIAAAMVLRGMAKDGQFALK--DAAAAHEGTVKNAVD 303
            | |   ||||||||||: ||||||||:||:|  |     || :| | :
Db    170 GADFNHEMKKDDQIAAAIALRGMAKDGKFAVKNDDEKGKAEGAIKGAAE 218


RESULT  14
O07056
ID   O07056      PRELIMINARY;      PRT;   209 AA.
AC   O07056;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   VMP-LIKE SEQUENCE PROTEIN VLSE (FRAGMENT).
GN   VLSE.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OG   Plasmid lp28-1.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B31-M1E4D;
RX   MEDLINE=97262068; PubMed=9108482;
RA   Zhang J.R., Hardham J.M., Barbour A.G., Norris S.J.;
RT   "Antigenic variation in Lyme disease borreliae by promiscuous
RT   recombination of VMP-like sequence cassettes.";
RL   Cell 89:275-285(1997).
DR   EMBL; U84557; AAC45194.1; -.
DR   InterPro; IPR000680; Borrelia_lipo.
DR   Pfam; PF00921; Lipoprotein_2; 1.
KW   Plasmid.
FT   NON_TER       1      1
FT   NON_TER     209    209
SQ   SEQUENCE   209 AA;  20214 MW;  5350AE7F83B7B38C CRC64;


  Query Match             24.8%;  Score 423.5;  DB 2;  Length 209;
  Best Local Similarity   49.3%;  Pred. No. 3.9e-13;
  Matches  111;  Conservative   24;  Mismatches   69;  Indels   21;  Gaps    7;

Qy      79 GKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGDVVNGDV 138
           || |:||: ||         |: |  :  :  |    |   |  |::|  |||:||: |
Db       4 GKPDSTGSVGTA--------VEGAIKEVSELLDKLVTAVKTAEGASSGTDAIGEVVDND- 54

Qy     139 AKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNAAGAEGTTNADAGKLFVKNAGNVG 198
             ||  | ||| |||||||| ||:||  ::    ||  | |:| :|  |||||  |   :
Db      55 --AKAADKASVTGIAKGIKEIVEAAGGSE----KLKVAAAKGESNKGAGKLFGKAGADAN 108

Qy     199 GEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAADRTNPIDAAIG-GAGDNDA 257
           |::  | ||| ||:||||||||| ||| || |  |: |:|  |  ||| |||| | || |
Db     109 GDSEAASKAAGAVSAVSGEQILSAIVTAA-DAAEQDGEKPGDAKNPIAAAIGKGDGDADF 167

Qy     258 AAAFATMKKDDQIAAAMVLRGMAKDGQFALK-DAAAAHEGTVKNA 301
                |||||||||: ||||||||:||:| |    || :| |
Db     168 GDG---MKKDDQIAAAIALRGMAKDGKFAVKNDEKGKAEGAIKGA 209


RESULT  15
O68358
ID   O68358      PRELIMINARY;      PRT;   214 AA.
AC   O68358;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   VMP-LIKE SEQUENCE PROTEIN VSLE (FRAGMENT).
GN   VLSE.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OG   Plasmid lp28-1.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B31-1380C;
RA   Zhang J.-R., Norris S.J.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF034506; AAC31335.1; -.
DR   InterPro; IPR000680; Borrelia_lipo.
DR   Pfam; PF00921; Lipoprotein_2; 1.
KW   Plasmid.
FT   NON_TER       1      1
FT   NON_TER     214    214
SQ   SEQUENCE   214 AA;  20857 MW;  12B38DEE1A457277 CRC64;


  Query Match             24.7%;  Score 422;  DB 2;  Length 214;
  Best Local Similarity   49.8%;  Pred. No. 4.7e-13;
  Matches  113;  Conservative   21;  Mismatches   73;  Indels   20;  Gaps    7;

Qy      79 GKLDATGAEGTTNVNAGKLFVKRAADDGGDADDAGKAAAAVAASAATGNAAIGDVVNGDV 138
           || |:||: ||         |: |  :  :  |    |   |  |::| ||||:||  |
Db       4 GKPDSTGSVGTA--------VEGAIKEVSELLDKLVKAVKTAEGASSGTAAIGEVVANDA 55

Qy     139 AKAKGGDAASVNGIAKGIKGIVDAAEKADAKEGKLNA-AGAEGTTNADAGKLFVKNAGNV 197
             ||  | ||| |||||||| ||:||  ::    || | | |:|  |  |||||
Db      56 DAAKVADKASVKGIAKGIKEIVEAAGGSE----KLKAVAAAKGENNKGAGKLFGKAGAAA 111

Qy     198 GGEAGDAGKAAAAVAAVSGEQILKAIVHAAKDGGEKQGKKAADRTNPIDAAIGGAGDNDA 257
           |::  | ||| ||:||||||||| ||| || |  |: |||  :  ||| |||   || |
Db     112 HGDSEAASKAAGAVSAVSGEQILSAIVTAA-DAAEQDGKKPEEAKNPIAAAI---GDKDG 167

Qy     258 AAAFA--TMKKDDQIAAAMVLRGMAKDGQFALKDAAAAH-EGTVKNA 301
           | |    ||||||||||: ||||||||:||:||      || :| |
Db     168 GAEFGQDEMKKDDQIAAAIALRGMAKDGKFAVKDGEKEKAEGAIKGA 214


Search completed: January 10, 2002, 14:09:34
Job time: 333 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2002, 14:09:34 ; Search time 38 Seconds
(without alignments)
180.916 Million cell updates/sec

Title: US-09-445-803-14
Perfect score: 235
Sequence: 1 PLTNPIDAAIGGSADRNAEA.....DQIAAAMVLRGMAKDGQFAL 47

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

T[redacted] number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_17:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

[redacted]

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	164	69.8	192	2	Q9RHW5	Q9rhw5 borrelia bu
2	164	69.8	223	2	Q9RHW7	Q9rhw7 borrelia bu
3	154	65.5	190	2	Q9RHW4	Q9rhw4 borrelia bu
4	149.5	63.6	211	2	O68379	O68379 borrelia bu
5	149.5	63.6	218	2	Q9RF39	Q9rf39 borrelia bu
6	148.5	63.2	192	2	O06881	O06881 borrelia bu
7	147.5	62.8	217	2	O07061	O07061 borrelia bu
8	145.5	61.9	358	2	O07055	O07055 borrelia bu
9	144.5	61.5	190	2	O06882	O06882 borrelia bu
10	144.5	61.5	211	2	O68351	O68351 borrelia bu
11	144.5	61.5	212	2	O68363	O68363 borrelia bu
12	144.5	61.5	213	2	O68354	O68354 borrelia bu
13	144.5	61.5	213	2	O07057	O07057 borrelia bu
14	144	61.3	189	2	O06888	O06888 borrelia bu
15	143.5	61.1	215	2	O68371	O68371 borrelia bu
16	143.5	61.1	216	2	O68349	O68349 borrelia bu
17	143.5	61.1	216	2	O68352	O68352 borrelia bu
18	143.5	61.1	356	2	O07053	O07053 borrelia bu
19	142.5	60.6	215	2	O07059	O07059 borrelia bu
20	142	60.4	169	2	O06879	O06879 borrelia bu
21	142	60.4	210	2	O07064	O07064 borrelia bu
22	141.5	60.2	186	2	O06886	O06886 borrelia bu
23	141.5	60.2	189	2	Q9RHW6	Q9rhw6 borrelia bu
24	141.5	60.2	210	2	O07054	O07054 borrelia bu
25	141.5	60.2	211	2	O68301	O68301 borrelia bu
26	141.5	60.2	212	2	O68341	O68341 borrelia bu
27	141.5	60.2	212	2	O68344	O68344 borrelia bu
28	141.5	60.2	212	2	O68360	O68360 borrelia bu
29	141.5	60.2	214	2	O68302	O68302 borrelia bu
30	141.5	60.2	218	2	O68353	O68353 borrelia bu
31	140.5	59.8	190	2	O06889	O06889 borrelia bu
32	140.5	59.8	212	2	O68361	O68361 borrelia bu
33	140.5	59.8	213	2	O68373	O68373 borrelia bu
34	140.5	59.8	214	2	O68362	O68362 borrelia bu
35	140.5	59.8	222	2	Q9RF44	Q9rf44 borrelia bu
36	140.5	59.8	223	2	Q9RF43	Q9rf43 borrelia bu
37	139.5	59.4	212	2	O68372	O68372 borrelia bu
38	139.5	59.4	216	2	O68355	O68355 borrelia bu
39	139	59.1	176	2	O06890	O06890 borrelia bu
40	139	59.1	215	2	O68364	O68364 borrelia bu
41	139	59.1	216	2	O68365	O68365 borrelia bu
42	139	59.1	217	2	O68304	O68304 borrelia bu
43	139	59.1	221	2	Q9RF38	Q9rf38 borrelia bu
44	138	58.7	211	2	O68305	O68305 borrelia bu
45	138	58.7	214	2	O07058	O07058 borrelia bu

ALIGNMENTS

```
RESULT   1
Q9RHW5
ID   Q9RHW5      PRELIMINARY;      PRT;   192 AA.
AC   Q9RHW5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   VLSF8 PROTEIN (FRAGMENT).
GN   VLSF8.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OG   Plasmid 20-kb borrelial plasmid.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=297;
RX   MEDLINE=98181050; PubMed=9514637;
RA   Kawabata H., Myouga F., Inagaki Y., Murai N., Watanabe H.;
RT   "Genetic and immunological analyses of Vls (VMP-like sequences) of
RT   Borrelia burgdorferi.";
RL   Microb. Pathog. 24:155-166(1998).
DR   EMBL; AB011065; BAA87888.1; -.
DR   InterPro; IPR000680; Borrelia_lipo.
DR   Pfam; PF00921; Lipoprotein_2; 1.
KW   Plasmid.
FT   NON_TER       1      1
FT   NON_TER     192    192
SQ   SEQUENCE   192 AA;  18638 MW;  F45DE1DF40A81A12 CRC64;


  Query Match             69.8%;  Score 164;  DB 2;  Length 192;
  Best Local Similarity   72.3%;  Pred. No. 2.3e-13;
  Matches   34;  Conservative    7;  Mismatches    4;  Indels    2;  Gaps    1;

Qy       3 TNPIDAAIGGSADRNAEAF--DKMKKDDQIAAAMVLRGMAKDGQFAL 47
           ||||:|||| : | :| ||  |:|||:||||||:||||||||:|||
Db     131 TNPIEAAIGSTGDNDAAAFSQDEMKKNDQIAAAIVLRGMAKDGEFAL 177


RESULT   2
Q9RHW7
ID   Q9RHW7      PRELIMINARY;      PRT;   223 AA.
```

```
AC   Q9RHW7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   VLSII PROTEIN (FRAGMENT).
GN   VLSII.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OG   Plasmid 20-kb borrelial plasmid.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=297;
RX   MEDLINE=98181050; PubMed=9514637;
RA   Kawabata H., Myouga F., Inagaki Y., Murai N., Watanabe H.;
RT   "Genetic and immunological analyses of Vls (VMP-like sequences) of
RT   Borrelia burgdorferi.";
RL   Microb. Pathog. 24:155-166(1998).
DR   EMBL; AB011063; BAA87886.1; -.
DR   InterPro; IPR000680; Borrelia_lipo.
DR   Pfam; PF00921; Lipoprotein_2; 1.
KW   Plasmid.
FT   NON_TER     223    223
SQ   SEQUENCE   223 AA;  22053 MW;  47080F3C3C45C82A CRC64;

  Query Match             69.8%;  Score 164;  DB 2;  Length 223;
  Best Local Similarity   72.3%;  Pred. No. 2.7e-13;
  Matches   34;  Conservative    7;  Mismatches    4;  Indels    2;  Gaps    1;

Qy       3 TNPIDAAIGGSADRNAEAF--DKMKKDDQIAAAMVLRGMAKDGQFAL 47
           ||||:|||| : | :| ||  |:|||:||||||:||||||||:|||
Db     142 TNPIEAAIGSTGDNDAAAFSQDEMKKNDQIAAAIVLRGMAKDGEFAL 188

RESULT   3
Q9RHW4
ID   Q9RHW4      PRELIMINARY;      PRT;   190 AA.
AC   Q9RHW4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   VLSF15 PROTEIN (FRAGMENT).
GN   VLSF15.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OG   Plasmid 20-kb borrelial plasmid.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=297;
RX   MEDLINE=98181050; PubMed=9514637;
RA   Kawabata H., Myouga F., Inagaki Y., Murai N., Watanabe H.;
RT   "Genetic and immunological analyses of Vls (VMP-like sequences) of
RT   Borrelia burgdorferi.";
RL   Microb. Pathog. 24:155-166(1998).
DR   EMBL; AB011066; BAA87889.1; -.
DR   InterPro; IPR000680; Borrelia_lipo.
DR   Pfam; PF00921; Lipoprotein_2; 1.
KW   Plasmid.
FT   NON_TER       1      1
FT   NON_TER     190    190
SQ   SEQUENCE   190 AA;  18748 MW;  DF5EA566DA6D29CB CRC64;

  Query Match             65.5%;  Score 154;  DB 2;  Length 190;
  Best Local Similarity   69.6%;  Pred. No. 4.3e-12;
  Matches   32;  Conservative    7;  Mismatches    5;  Indels    2;  Gaps    1;

Qy       4 NPIDAAIGGSADRNAEAF--DKMKKDDQIAAAMVLRGMAKDGQFAL 47
           ||| |||| : | :| ||  |:|||:||:|||:||||||||:|||
Db     130 NPIAAAIGSTGDNDAAAFSKDEMKKNDQVAAAIVLRGMAKDGEFAL 175

RESULT   4
O68379
ID   O68379      PRELIMINARY;      PRT;   211 AA.
AC   O68379;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   VMP-LIKE SEQUENCE PROTEIN VSLE (FRAGMENT).
GN   VLSE.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OG   Plasmid lp28-1.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B31-1502E;
RA   Zhang J.-R., Norris S.J.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF034527; AAC31356.1; -.
DR   InterPro; IPR000680; Borrelia_lipo.
DR   Pfam; PF00921; Lipoprotein_2; 1.
KW   Plasmid.
FT   NON_TER       1      1
FT   NON_TER     211    211
SQ   SEQUENCE   211 AA;  20580 MW;  57420385FD714C35 CRC64;

  Query Match             63.6%;  Score 149.5;  DB 2;  Length 211;
  Best Local Similarity   71.7%;  Pred. No. 1.9e-11;
  Matches   33;  Conservative    5;  Mismatches    7;  Indels    1;  Gaps    1;

Qy       3 TNPIDAAIG-GSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
           |||| |||| | ||  ||  ::||||||||||: |||||||:||:
Db     151 TNPIAAAIGKGDADDGAEFGNEMKKDDQIAAAIALRGMAKDGKFAV 196

RESULT   5
Q9RF39
ID   Q9RF39      PRELIMINARY;      PRT;   218 AA.
AC   Q9RF39;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   VLSE (FRAGMENT).
GN   VLSE.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W2F;
RA   Iyer R., Hardham J.M., Wormser G.P., Schwartz I., Norris S.J.;
RT   "Conservation and heterogeneity of vlsE among human and tick isolates
RT   of Borrelia burgdorferi.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF201680; AAF25656.1; -.
DR   InterPro; IPR000680; Borrelia_lipo.
DR   Pfam; PF00921; Lipoprotein_2; 1.
FT   NON_TER       1      1
FT   NON_TER     218    218
SQ   SEQUENCE   218 AA;  21052 MW;  C850839CEDA0E9E4 CRC64;

  Query Match             63.6%;  Score 149.5;  DB 2;  Length 218;
  Best Local Similarity   71.7%;  Pred. No. 1.9e-11;
  Matches   33;  Conservative    5;  Mismatches    7;  Indels    1;  Gaps    1;

Qy       3 TNPIDAAIG-GSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
           |||| |||| |:||  |:  | ||||||||||: ||||||||:||:
Db     150 TNPIAAAIGKGNADDGADFGDGMKKDDQIAAAIALRGMAKDGKFAV 195
```

```
RESULT   6
O06881
ID   O06881      PRELIMINARY;      PRT;   192 AA.
AC   O06881;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PUTATIVE VLS RECOMBINATION CASSETTE VLS4 (FRAGMENT).
OS   Borrelia burgdorferi (Lyme disease spirochete).
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B31 (ATCC 35210) CLONE 5A3;
RX   MEDLINE=97262068; PubMed=9108482;
RA   Zhang J.R., Hardham J.M., Barbour A.G., Norris S.J.;
RT   "Antigenic variation in Lyme disease borreliae by promiscuous
RT   recombination of VMP-like sequence cassettes.";
RL   Cell 89:275-285(1997).
DR   EMBL; U76406; AAC45178.1; -.
DR   InterPro; IPR000680; Borrelia_lipo.
DR   Pfam; PF00921; Lipoprotein_2; 1.
FT   NON_TER       1      1
FT   NON_TER     192    192
SQ   SEQUENCE   192 AA;  18519 MW;  78A5A2257F7844E8 CRC64;


  Query Match             63.2%;  Score 148.5;  DB 2;  Length 192;
  Best Local Similarity   69.6%;  Pred. No. 2.2e-11;
  Matches   32;  Conservative    6;  Mismatches    7;  Indels    1;  Gaps    1;

Qy       3 TNPIDAAIG-GSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
           |||| |||| |: :  ||  |:||||||||||: ||||||||:||:
Db     138 TNPIAAAIGKGNEENGAEFKDEMKKDDQIAAAIALRGMAKDGKFAV 183


RESULT   7
O07061
ID   O07061      PRELIMINARY;      PRT;   217 AA.
AC   O07061;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   VMP-LIKE SEQUENCE PROTEIN VLSE (FRAGMENT).
GN   VLSE.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OG   Plasmid lp28-1.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B31-M4B4A;
RX   MEDLINE=97262068; PubMed=9108482;
RA   Zhang J.R., Hardham J.M., Barbour A.G., Norris S.J.;
RT   "Antigenic variation in Lyme disease borreliae by promiscuous
RT   recombination of VMP-like sequence cassettes.";
RL   Cell 89:275-285(1997).
DR   EMBL; U84562; AAC45199.1; -.
DR   InterPro; IPR000680; Borrelia_lipo.
DR   Pfam; PF00921; Lipoprotein_2; 1.
KW   Plasmid.
FT   NON_TER       1      1
FT   NON_TER     217    217
SQ   SEQUENCE   217 AA;  20966 MW;  619E8CA8B3A97449 CRC64;


  Query Match             62.8%;  Score 147.5;  DB 2;  Length 217;
  Best Local Similarity   73.3%;  Pred. No. 3.5e-11;
  Matches   33;  Conservative    4;  Mismatches    7;  Indels    1;  Gaps    1;

Qy       4 NPIDAAIG-GSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
           ||| |||| |:||  ||  | ||||||||||: ||||||||:||:
Db     157 NPIAAAIGKGNADDGAEFGDGMKKDDQIAAAIALRGMAKDGKFAV 201


RESULT   8
O07055
ID   O07055      PRELIMINARY;      PRT;   358 AA.
AC   O07055;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   VMP-LIKE SEQUENCE PROTEIN VLSE.
GN   VLSE.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OG   Plasmid lp28-1.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B31-M1E4C;
RX   MEDLINE=97262068; PubMed=9108482;
RA   Zhang J.R., Hardham J.M., Barbour A.G., Norris S.J.;
RT   "Antigenic variation in Lyme disease borreliae by promiscuous
RT   recombination of VMP-like sequence cassettes.";
RL   Cell 89:275-285(1997).
DR   EMBL; U84556; AAC45764.1; -.
DR   InterPro; IPR000680; Borrelia_lipo.
DR   Pfam; PF00921; Lipoprotein_2; 1.
KW   Plasmid.
SQ   SEQUENCE   358 AA;  36009 MW;  6D22AE9AAA076EEC CRC64;


  Query Match             61.9%;  Score 145.5;  DB 2;  Length 358;
  Best Local Similarity   69.6%;  Pred. No. 1.1e-10;
  Matches   32;  Conservative    5;  Mismatches    8;  Indels    1;  Gaps    1;

Qy       3 TNPIDAAIG-GSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
           |||| |||| |: :  ||  | ||||||||||: ||||||||:||:
Db     253 TNPIAAAIGKGNEENGAEFGDGMKKDDQIAAAIALRGMAKDGKFAV 298


RESULT   9
O06882
ID   O06882      PRELIMINARY;      PRT;   190 AA.
AC   O06882;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PUTATIVE VLS RECOMBINATION CASSETTE VLS5 (FRAGMENT).
OS   Borrelia burgdorferi (Lyme disease spirochete).
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B31 (ATCC 35210) CLONE 5A3;
RX   MEDLINE=97262068; PubMed=9108482;
RA   Zhang J.R., Hardham J.M., Barbour A.G., Norris S.J.;
RT   "Antigenic variation in Lyme disease borreliae by promiscuous
RT   recombination of VMP-like sequence cassettes.";
RL   Cell 89:275-285(1997).
DR   EMBL; U76406; AAC45179.1; -.
DR   InterPro; IPR000680; Borrelia_lipo.
DR   Pfam; PF00921; Lipoprotein_2; 1.
FT   NON_TER       1      1
FT   NON_TER     190    190
SQ   SEQUENCE   190 AA;  18376 MW;  898EFF84308EAA7A CRC64;


  Query Match             61.5%;  Score 144.5;  DB 2;  Length 190;
  Best Local Similarity   71.1%;  Pred. No. 7.2e-11;
  Matches   32;  Conservative    5;  Mismatches    7;  Indels    1;  Gaps    1;

Qy       4 NPIDAAIG-GSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
           ||| |||| |:||  |:  | ||||||||||: ||||||||:||:
Db     138 NPIAAAIGKGNADDGADFGDGMKKDDQIAAAIALRGMAKDGKFAV 182
```

QH573.C38

```
RESULT  10
O68351
ID   O68351      PRELIMINARY;      PRT;   211 AA.
AC   O68351;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   VMP-LIKE SEQUENCE PROTEIN VSLE (FRAGMENT).
GN   VLSE.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OG   Plasmid lp28-1.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B31-1373D;
RA   Zhang J.-R., Norris S.J.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF034498; AAC31328.1; -.
DR   InterPro; IPR000680; Borrelia_lipo.
DR   Pfam; PF00921; Lipoprotein_2; 1.
KW   Plasmid.
FT   NON_TER       1      1
FT   NON_TER     211    211
SQ   SEQUENCE   211 AA;  20495 MW;  8DE5C9D42CF4526F CRC64;


  Query Match             61.5%;  Score 144.5;  DB 2;  Length 211;
  Best Local Similarity   71.1%;  Pred. No. 8.1e-11;
  Matches   32;  Conservative    5;  Mismatches    7;  Indels    1;  Gaps    1;

Qy       4 NPIDAAIG-GSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
           ||| |||| |:|| |:  | |||||||||| |||||||:||:
Db     152 NPIAAAIGKGNADDGADFGDGMKKDDQIAAAIALRGMAKDGKFAV 196


RESULT  11
O68363
ID   O68363      PRELIMINARY;      PRT;   212 AA.
AC   O68363;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   VMP-LIKE SEQUENCE PROTEIN VSLE (FRAGMENT).
GN   VLSE.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OG   Plasmid lp28-1.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B31-1394F;
RA   Zhang J.-R., Norris S.J.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF034511; AAC31340.1; -.
DR   InterPro; IPR000680; Borrelia_lipo.
DR   Pfam; PF00921; Lipoprotein_2; 1.
KW   Plasmid.
FT   NON_TER       1      1
FT   NON_TER     212    212
SQ   SEQUENCE   212 AA;  20568 MW;  493F7B407DE895F6 CRC64;


  Query Match             61.5%;  Score 144.5;  DB 2;  Length 212;
  Best Local Similarity   71.1%;  Pred. No. 8.2e-11;
  Matches   32;  Conservative    5;  Mismatches    7;  Indels    1;  Gaps    1;

Qy       4 NPIDAAIG-GSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
           ||| |||| |:|| |:  | |||||||||| |||||||:||:
Db     153 NPIAAAIGKGNADDGADFGDGMKKDDQIAAAIALRGMAKDGKFAV 197
```

```
RESULT  12
O68354
ID   O68354      PRELIMINARY;      PRT;   213 AA.
AC   O68354;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   VMP-LIKE SEQUENCE PROTEIN VSLE (FRAGMENT).
GN   VLSE.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OG   Plasmid lp28-1.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B31-1375A;
RA   Zhang J.-R., Norris S.J.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF034501; AAC31331.1; -.
DR   InterPro; IPR000680; Borrelia_lipo.
DR   Pfam; PF00921; Lipoprotein_2; 1.
KW   Plasmid.
FT   NON_TER       1      1
FT   NON_TER     213    213
SQ   SEQUENCE   213 AA;  20493 MW;  BEB670LC32DBF29E CRC64;


  Query Match             61.5%;  Score 144.5;  DB 2;  Length 213;
  Best Local Similarity   71.1%;  Pred. No. 8.2e-11;
  Matches   32;  Conservative    5;  Mismatches    7;  Indels    1;  Gaps    1;

Qy       4 NPIDAAIG-GSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
           ||| |||| |:|| |:  | |||||||||| |||||||:||:
Db     154 NPIAAAIGEGNADDGADFGDGMKKDDQIAAAIALRGMAKDGKFAV 198


RESULT  13
O07057
ID   O07057      PRELIMINARY;      PRT;   213 AA.
AC   O07057;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   VMP-LIKE SEQUENCE PROTEIN VLSE (FRAGMENT).
GN   VLSE.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OG   Plasmid lp28-1.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B31-M1E4E;
RX   MEDLINE=97262068; PubMed=9108482;
RA   Zhang J.R., Hardham J.M., Barbour A.G., Norris S.J.;
RT   "Antigenic variation in Lyme disease borreliae by promiscuous
RT   recombination of VMP-like sequence cassettes.";
RL   Cell 89:275-285(1997).
DR   EMBL; U84558; AAC45195.1; -.
DR   InterPro; IPR000680; Borrelia_lipo.
DR   Pfam; PF00921; Lipoprotein_2; 1.
KW   Plasmid.
FT   NON_TER       1      1
FT   NON_TER     213    213
SQ   SEQUENCE   213 AA;  20611 MW;  B98DC82F3D91645A CRC64;


  Query Match             61.5%;  Score 144.5;  DB 2;  Length 213;
  Best Local Similarity   71.1%;  Pred. No. 8.2e-11;
  Matches   32;  Conservative    5;  Mismatches    7;  Indels    1;  Gaps    1;

Qy       4 NPIDAAIG-GSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
           ||| |||| |:|| |:  | |||||||||| |||||||:||:
```

```
Db    154 NPIAAAIGKGNADDGADFGDGMKKDDQIAAAIALRGMAKDGKFAV 198


RESULT  14
O06888
ID   O06888      PRELIMINARY;      PRT;   189 AA.
AC   O06888;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PUTATIVE VLS RECOMBINATION CASSETTE VLS12 (FRAGMENT).
OS   Borrelia burgdorferi (Lyme disease spirochete).
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B31 (ATCC 35210) CLONE 5A3;
RX   MEDLINE=97262068; PubMed=9108482;
RA   Zhang J.R., Hardham J.M., Barbour A.G., Norris S.J.;
RT   "Antigenic variation in Lyme disease borreliae by promiscuous
RT   recombination of VMP-like sequence cassettes.";
RL   Cell 89:275-285(1997).
DR   EMBL; U76406; AAC45185.1; -.
DR   InterPro; IPR000680; Borrelia_lipo.
DR   Pfam; PF00921; Lipoprotein_2; 1.
FT   NON_TER       1      1
FT   NON_TER     189    189
SQ   SEQUENCE   189 AA;  18642 MW;  384F0D4BFD70422B CRC64;


  Query Match             61.3%;  Score 144;  DB 2;  Length 189;
  Best Local Similarity   68.2%;  Pred. No. 8.3e-11;
  Matches   30;  Conservative    5;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       4 NPIDAAIGGSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
           ||| ||||   :  ||  |:||||||||||: ||||||||:||:
Db     136 NPIAAAIGKGNENGAEFKDEMKKDDQIAAAIALRGMAKDGKFAV 179


RESULT  15
O68371
ID   O68371      PRELIMINARY;      PRT;   215 AA.
AC   O68371;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   VMP-LIKE SEQUENCE PROTEIN VSLE (FRAGMENT).
GN   VLSE.
OS   Borrelia burgdorferi (Lyme disease spirochete).
     Plasmid lp28-1.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B31-1413A;
RA   Zhang J.-R., Norris S.J.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF034519; AAC31348.1; -.
DR   HSSP; P48781; 1B67.
DR   InterPro; IPR000680; Borrelia_lipo.
DR   Pfam; PF00921; Lipoprotein_2; 1.
KW   Plasmid.
FT   NON_TER       1      1
FT   NON_TER     215    215
SQ   SEQUENCE   215 AA;  20803 MW;  973CC70DB9348542 CRC64;


  Query Match             61.1%;  Score 143.5;  DB 2;  Length 215;
  Best Local Similarity   68.9%;  Pred. No. 1.1e-10;
  Matches   31;  Conservative    6;  Mismatches    7;  Indels    1;  Gaps    1;

Qy       4 NPIDAAIG-GSADRNAEAFDKMKKDDQIAAAMVLRGMAKDGQFAL 47
           ||| |||| |: :  ||  |:||||||||||: ||||||||:||:

Db     156 NPIAAAIGKGNEENGAEFKDEMKKDDQIAAAIALRGMAKDGKFAV 200


Search completed: January 10, 2002, 14:09:35
Job time: 334 sec
```